GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:11:58 ; Search time 12.48 Seconds
(without alignments)
200.149 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 552
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	552	100.0	111	1	US-07-662-193-4	Sequence 4, Appli
2	552	100.0	111	3	US-08-300-928C-6	Sequence 6, Appli
3	552	100.0	111	3	US-08-430-944D-6	Sequence 6, Appli
4	552	100.0	111	3	US-08-430-014-6	Sequence 6, Appli
5	552	100.0	111	3	US-08-431-184-6	Sequence 6, Appli
6	541	98.0	109	1	US-07-807-529A-6	Sequence 6, Appli
7	541	98.0	109	5	PCT-US93-02462-6	Sequence 6, Appli
8	506	91.7	109	1	US-07-662-193-5	Sequence 5, Appli
9	506	91.7	109	3	US-08-300-928C-8	Sequence 8, Appli
10	506	91.7	109	3	US-08-430-944D-8	Sequence 8, Appli
11	506	91.7	109	3	US-08-430-014-8	Sequence 8, Appli
12	506	91.7	109	3	US-08-431-184-8	Sequence 8, Appli
13	465	84.2	92	3	US-08-300-928C-13	Sequence 13, Appl
14	465	84.2	92	3	US-08-430-944D-13	Sequence 13, Appl
15	465	84.2	92	3	US-08-430-014-13	Sequence 13, Appl
16	465	84.2	92	3	US-08-431-184-13	Sequence 13, Appl
17	419	75.9	90	3	US-08-300-928C-14	Sequence 14, Appl
18	419	75.9	90	3	US-08-430-944D-14	Sequence 14, Appl
19	419	75.9	90	3	US-08-430-014-14	Sequence 14, Appl
20	419	75.9	90	3	US-08-431-184-14	Sequence 14, Appl
21	409	74.1	83	1	US-07-662-193-7	Sequence 7, Appli
22	408	73.9	97	3	US-08-300-928C-10	Sequence 10, Appl
23	408	73.9	97	3	US-08-430-944D-10	Sequence 10, Appl
24	408	73.9	97	3	US-08-430-014-10	Sequence 10, Appl
25	408	73.9	97	3	US-08-431-184-10	Sequence 10, Appl
26	383	69.4	83	1	US-07-662-193-8	Sequence 8, Appli
27	370.5	67.1	82	3	US-08-300-928C-16	Sequence 16, Appl
28	370.5	67.1	82	3	US-08-430-944D-16	Sequence 16, Appl
29	370.5	67.1	82	3	US-08-430-014-16	Sequence 16, Appl
30	370.5	67.1	82	3	US-08-431-184-16	Sequence 16, Appl
31	185	33.5	51	1	US-07-662-193-6	Sequence 6, Appli
32	185	33.5	51	3	US-08-300-928C-15	Sequence 15, Appl
33	185	33.5	51	3	US-08-430-944D-15	Sequence 15, Appl
34	185	33.5	51	3	US-08-430-014-15	Sequence 15, Appl
35	185	33.5	51	3	US-08-431-184-15	Sequence 15, Appl
36	128.5	23.3	96	1	US-07-807-529A-39	Sequence 39, Appl
37	128.5	23.3	96	3	US-08-430-944D-103	Sequence 103, App
38	128.5	23.3	96	3	US-08-431-184-103	Sequence 103, App
39	128	23.2	26	1	US-07-807-529A-9	Sequence 9, Appli
40	128	23.2	26	3	US-08-300-928C-19	Sequence 19, Appl
41	128	23.2	26	3	US-08-430-944D-19	Sequence 19, Appl
42	128	23.2	26	3	US-08-430-014-19	Sequence 19, Appl
43	128	23.2	26	3	US-08-431-184-19	Sequence 19, Appl
44	128	23.2	26	5	PCT-US93-02462-9	Sequence 9, Appli
45	118.5	21.5	27	3	US-08-300-928C-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-07-662-193-4
; Sequence 4, Application US/07662193
; Patent No. 5328991
; GENERAL INFORMATION:
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Bond, Julian
; TITLE OF INVENTION: Improved Preparation of Cat Dander
; TITLE OF INVENTION: Allergens for Immunotherapeutic Purposes and Uses Thereof
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/662,193
; FILING DATE: 19910228
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,276
; FILING DATE: 28-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: P36,207
; REFERENCE/DOCKET NUMBER: IML89-02AA/IPC-002CC/IMI-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-07-662-193-4

Query Match 100.0%; Score 552; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 5.7e-64;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   2
US-08-300-928C-6
; Sequence 6, Application US/08300928C
; Patent No. 6019972
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,928C
;      FILING DATE:  September 2, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,529
;      FILING DATE:  December 13, 1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  111 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-300-928C-6


  Query Match             100.0%;  Score 552;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-64;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   3
US-08-430-944D-6
; Sequence 6, Application US/08430944D
; Patent No. 6025162
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,944D
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  111 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-430-944D-6


  Query Match             100.0%;  Score 552;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-64;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   4
US-08-430-014-6
; Sequence 6, Application US/08430014
; Patent No. 6048962
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/430,014
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/300,928
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  AMY E. MANDRAGOURAS
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  111 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-430-014-6


  Query Match             100.0%;  Score 552;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-64;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   5
US-08-431-184-6
; Sequence 6, Application US/08431184
; Patent No. 6120769
;  GENERAL INFORMATION:
;     APPLICANT:  Bruce L. Rogers et al.
;     TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;     TITLE OF INVENTION:
;     NUMBER OF SEQUENCES:  103
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  28 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/431,184
;       FILING DATE:  28-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/430,014
;       FILING DATE:  27-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/300,928
;       FILING DATE:  02-SEPT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amy E. Mandragouras
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  IMI-044DV3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)742-4214
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  111 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-431-184-6


  Query Match             100.0%;  Score 552;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-64;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   6
US-07-807-529A-6
; Sequence 6, Application US/07807529A
; Patent No. 5547669
;  GENERAL INFORMATION:
;     APPLICANT:  Rogers, Bruce L.
;     APPLICANT:  Morgenstern, Jay
;     APPLICANT:  Bond, Julian F.
;     APPLICANT:  Garman, Richard D.
;     APPLICANT:  Greenstein, Julia L.
;     APPLICANT:  Kuo, Mei-chang
;     APPLICANT:  Morvile, Malcolm
;     TITLE OF INVENTION:  RECOMBITOPE PEPTIDES
;     NUMBER OF SEQUENCES:  76
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;       STREET:  One Kendall Square, Building 600
;       CITY:  Cambridge
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02139
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII TEXT
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/807,529A
;       FILING DATE:  19911213
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/662,276
;       FILING DATE:  28-FEB-1991
;       APPLICATION NUMBER:  US 07/431,565
;       FILING DATE:  03-NOV-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Channing, Stacey L.
;       REGISTRATION NUMBER:  31,095
;       REFERENCE/DOCKET NUMBER:  IPC-027/imi-015
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 494-0060
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  109 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-807-529A-6
```

```
  Query Match             98.0%;  Score 541;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-62;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109


RESULT   7
PCT-US93-02462-6
; Sequence 6, Application PC/TUS9302462
;  GENERAL INFORMATION:
;    APPLICANT:  Gefter, Malcolm L.
;    APPLICANT:  Garman, Richard D.
;    APPLICANT:  Greenstein, Julia L.
;    APPLICANT:  Kuo, Mei-chang
;    APPLICANT:  Briner, Thomas J.
;    APPLICANT:  Morville, Malcolm
;    TITLE OF INVENTION:  PEPTIDES USEFUL FOR TOLERIZATION
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02462
;      FILING DATE:  19930325
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/006,116
;      FILING DATE:  15-JAN-1993
;      APPLICATION NUMBER:  US 07/884,718
;      FILING DATE:  15-MAY-1992
;      APPLICATION NUMBER:  07/857,311
;      FILING DATE:  25-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-031PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-02462-6


  Query Match             98.0%;  Score 541;  DB 5;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-62;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109


RESULT   8
US-07-662-193-5
; Sequence 5, Application US/07662193
; Patent No. 5328991
;  GENERAL INFORMATION:
;    APPLICANT:  Kuo, Mei-chang
;    APPLICANT:  Bond, Julian
;    TITLE OF INVENTION:  Improved Preparation of Cat Dander
;    TITLE OF INVENTION:  Allergens for Immunotherapeutic Purposes and Uses Thereof
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/662,193
;      FILING DATE:  19910228
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/662,276
;      FILING DATE:  28-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  P36,207
;      REFERENCE/DOCKET NUMBER:  IML89-02AA/IPC-002CC/IMI-020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-07-662-193-5


  Query Match             91.7%;  Score 506;  DB 1;  Length 109;
  Best Local Similarity   93.7%;  Pred. No. 4.9e-58;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM--IAINEYCMGEAVQNTVEDLKLNTLGR 109


RESULT   9
US-08-300-928C-8
; Sequence 8, Application US/08300928C
; Patent No. 6019972
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;       STREET:  610 LINCOLN STREET
;       CITY:  WALTHAM
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02145
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/300,928C
;       FILING DATE:  September 2, 1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/807,529
;       FILING DATE:  December 13, 1991
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  AMY E. MANDRAGOURAS
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  109 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-300-928C-8


  Query Match             91.7%;  Score 506;  DB 3;  Length 109;
  Best Local Similarity   93.7%;  Pred. No. 4.9e-58;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM--IAINEYCMGEAVQNTVEDLKLNTLGR 109


RESULT  10
US-08-430-944D-8
; Sequence 8, Application US/08430944D
; Patent No. 6025162
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,944D
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-430-944D-8


  Query Match             91.7%;  Score 506;  DB 3;  Length 109;
  Best Local Similarity   93.7%;  Pred. No. 4.9e-58;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM--IAINEYCMGEAVQNTVEDLKLNTLGR 109


RESULT  11
US-08-430-014-8
; Sequence 8, Application US/08430014
; Patent No. 6048962
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,014
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/300,928
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
```

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-430-014-8


  Query Match             91.7%;  Score 506;  DB 3;  Length 109;
  Best Local Similarity   93.7%;  Pred. No. 4.9e-58;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM--IAINEYCMGEAVQNTVEDLKLNTLGR 109


RESULT  12
US-08-431-184-8
; Sequence 8, Application US/08431184
; Patent No. 6120769
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/431,184
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-431-184-8


  Query Match             91.7%;  Score 506;  DB 3;  Length 109;
  Best Local Similarity   93.7%;  Pred. No. 4.9e-58;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM--IAINEYCMGEAVQNTVEDLKLNTLGR 109


RESULT  13
US-08-300-928C-13
; Sequence 13, Application US/08300928C
; Patent No. 6019972
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,928C
;      FILING DATE:  September 2, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,529
;      FILING DATE:  December 13, 1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-300-928C-13


  Query Match             84.2%;  Score 465;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-53;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


RESULT  14
US-08-430-944D-13
; Sequence 13, Application US/08430944D
; Patent No. 6025162
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,944D
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-430-944D-13


  Query Match             84.2%;  Score 465;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-53;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


RESULT  15
US-08-430-014-13
; Sequence 13, Application US/08430014
; Patent No. 6048962
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,014
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/300,928
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-430-014-13


  Query Match             84.2%;  Score 465;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-53;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


Search completed: May  7, 2002, 12:13:47
Job time: 109 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:12:13 ; Search time 14.32 Seconds
(without alignments)
590.459 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 552
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	541	98.0	109	2	C56413	major allergen Fel
2	495	89.7	107	2	JC1127	major allergen cha
3	88	15.9	20	2	B53283	major cat allergen
4	76.5	13.9	2401	2	T28676	rhoptry protein -
5	73	13.2	1104	2	S36773	GTPase-activating
6	72	13.0	521	1	T17196	IMP dehydrogenase
7	72	13.0	800	2	D86712	glycogen phosphory
8	71	12.9	853	2	A28668	DNA mismatch repai
9	71	12.9	1252	1	S77037	hypothetical prote
10	70.5	12.8	295	2	A72219	conserved hypothet
11	70.5	12.8	388	2	T00641	hypothetical prote
12	70	12.7	245	2	D82610	ribose-5-phosphate
13	69	12.5	853	2	I54964	DNA mismatch repai
14	69	12.5	853	2	F85922	methyl-directed mi
15	69	12.5	857	2	T37459	ribonucleotide red
16	69	12.5	895	2	D72571	probable DNA-direc
17	68.5	12.4	466	2	H71234	probable 4-aminobu
18	68.5	12.4	534	2	F96713	unknown protein T6
19	68.5	12.4	1167	2	T06146	disease resistance
20	68	12.3	520	1	S22399	leucyl aminopeptid
21	67.5	12.2	346	2	C85088	peroxidase C2 prec
22	67	12.1	587	2	A34084	extracellular prot
23	67	12.1	1343	1	H64073	DNA-directed RNA p
24	66.5	12.0	955	2	T39765	probable nuclear m
25	65.5	11.9	351	2	S00626	peroxidase (EC 1.1
26	65.5	11.9	352	2	T05723	peroxidase (EC 1.1
27	65.5	11.9	468	2	H70746	probable hemA prot
28	65.5	11.9	1302	2	T00038	hypothetical prote
29	65.5	11.9	5825	2	T12117	polyprotein - fava
30	65	11.8	326	2	G86462	AIG1-like protein,
31	65	11.8	392	2	B81187	AcrA/AcrE family p
32	65	11.8	392	2	B81916	probable periplasm
33	65	11.8	522	2	B64067	Na+/H+-exchanging
34	65	11.8	530	2	C82144	Na+/H+ antiporter
35	65	11.8	870	2	G69006	DNA-directed RNA p
36	65	11.8	2412	1	JQ1537	genome polyprotein
37	65	11.8	4128	2	JC6306	protein kinase (EC
38	64.5	11.7	114	2	A55010	neutrophil-activat
39	64.5	11.7	229	2	F83663	serine O-acetyltra
40	64.5	11.7	833	2	T28385	ORF MSV224 probabl
41	64.5	11.7	936	2	H71728	2-oxoglutarate deh
42	64.5	11.7	1116	2	B70476	hypothetical prote
43	64.5	11.7	2526	2	T20531	hypothetical prote
44	64.5	11.7	2722	2	T20532	hypothetical prote
45	64	11.6	138	2	T07353	ribosomal protein

ALIGNMENTS

RESULT 1
C56413
major allergen Fel dI chain 2 precursor - cat
C;Species: Felis silvestris catus (domestic cat)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 01-Dec-2000
C;Accession: C56413; JC1145
R;Morgenstern, J.P.; Griffith, I.J.; Brauer, A.W.; Rogers, B.L.; Bond, J.F.; Chapman,
Proc. Natl. Acad. Sci. U.S.A. 88, 9690-9694, 1991
A;Title: Amino acid sequence of Fel dI, the major allergen of the domestic cat: prote
A;Reference number: A56413; MUID:92052157
A;Accession: C56413
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-109 <MOR>
A;Cross-references: GB:M77341; NID:g163822; PIDN:AAC41616.1; PID:g163823
R;Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.
Gene 113, 263-268, 1992
A;Title: Expression and genomic structure of the genes encoding FdI, the major allerg
A;Reference number: JC1126; MUID:92241678
A;Accession: JC1145
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 18-109 <GRI>
A;Experimental source: salivary gland
C;Keywords: glycoprotein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 98.0%; Score 541; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 7.1e-50;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109
```

RESULT 2
JC1127
major allergen chain 2 precursor, short form - cat
C;Species: Felis silvestris catus (domestic cat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: JC1127
R;Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.
Gene 113, 263-268, 1992
A;Title: Expression and genomic structure of the genes encoding FdI, the major allerg
A;Reference number: JC1126; MUID:92241678

```
A;Accession: JC1127
A;Molecule type: DNA
A;Residues: 1-107 <GRI>
A;Cross-references: GB:X62478; NID:g395406; PIDN:CAA44345.1; PID:g395407
A;Experimental source: skin
C;Genetics:
A;Gene: Ch2
A;Introns: 21/1; 81/3
C;Keywords: glycoprotein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-107/Product: major allergen chain 2, short form #status predicted <MAT>
F;50/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             89.7%;  Score 495;  DB 2;  Length 107;
  Best Local Similarity   93.6%;  Pred. No. 4.8e-45;
  Matches  102;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVM--IAINEYCMGEAVQNTVEDLKLNTLGR 107


RESULT   3
B53283
major cat allergen Fel d I beta chain - cat (fragment)
C;Species: Felis silvestris catus (domestic cat)
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 12-May-1994
C;Accession: B53283
R;Duffort, O.A.; Carreira, J.; Nitti, G.; Polo, F.; Lombardero, M.
Mol. Immunol. 28, 301-309, 1991
A;Title: Studies on the biochemical structure of the major cat allergen Felis domesticus
A;Reference number: A53283; MUID:91287714
A;Accession: B53283
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <DUF>


  Query Match             15.9%;  Score 88;  DB 2;  Length 20;
  Best Local Similarity   90.0%;  Pred. No. 0.0044;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANG 39
           |||||| ||||||| |||||
Db       1 VKMAETFPIFYDVFTAVANG 20


RESULT   4
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28676; A45521
R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A;Title: Comparison of two members of a multigene family coding for high-molecular mass
A;Reference number: Z20507; MUID:97077455
A;Accession: T28676
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2401 <SIN>
A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cc
A;Reference number: A45521; MUID:91101660
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEE>
A;Cross-references: GB:M34281


  Query Match             13.9%;  Score 76.5;  DB 2;  Length 2401;
  Best Local Similarity   35.1%;  Pred. No. 19;
  Matches   27;  Conservative   12;  Mismatches   25;  Indels   13;  Gaps    4;

Qy      34 FAVANGNELLLDLSLTKV-NATEPERTAMKKIQDCYVE-NGLISRVLDGLVMTTISSSKD 91
           |:  | | |     | :: ||   |   :|||:|  :| ||||:::        | : ||
Db    1732 FSEENNNILQSKKKLKELTNAFNAE---IKKIEDKIIEKNGLINKL--------IETRKD 1780

Qy      92 CMGEAVQNTVEDLKLNT 108
           ||     :  || ||: |
Db    1781 CMLFTYKTLVETLKIKT 1797


RESULT   5
S36773
GTPase-activating protein BUD2 - yeast (Saccharomyces cerevisiae)
N;Alternate names: cyclin CLA2; protein YKL092c; protein YKL424
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999
C;Accession: S36773; S37917; S37918; S41684; S39117; S42005; S34706
R;Park, H.O.; Chant, J.; Herskowitz, I.
Nature 365, 269-274, 1993
A;Title: BUD2 encodes a GTPase-activating protein for Bud1/Rsr1 necessary for proper
A;Reference number: S36773; MUID:93382538
A;Accession: S36773
A;Molecule type: DNA
A;Residues: 1-1104 <PAR>
A;Cross-references: EMBL:L19162; NID:g402733; PIDN:AAA34461.1; PID:g402734
R;Pohl, T.M.; Pohl, F.M.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37897
A;Accession: S37917
A;Molecule type: DNA
A;Residues: 1-681 <POH>
A;Cross-references: EMBL:Z28092; MIPS:YKL092c
A;Experimental source: strain S288C
R;James, C.M.; Gent, M.E.; Indge, K.J.; Oliver, S.G.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37918
A;Accession: S37918
A;Molecule type: DNA
A;Residues: 237-1104 <JAM>
A;Cross-references: EMBL:Z28092; MIPS:YKL092c
A;Experimental source: strain S288C
R;Cvrckova, F.; Nasmyth, K.
EMBO J. 12, 5277-5286, 1993
A;Title: Yeast G(1) cyclins CLN1 and CLN2 and a GAP-like protein have a role in bud f
A;Reference number: S41683; MUID:94085403
A;Accession: S41684
A;Molecule type: DNA
A;Residues: 1-436,'Y',438-1104 <CVR>
A;Cross-references: EMBL:X74130; NID:g395287; PIDN:CAA52228.1; PID:g395289
R;James, C.M.
submitted to the EMBL Data Library, October 1993
A;Reference number: S39116
A;Accession: S39117
A;Molecule type: DNA
A;Residues: 237-1104 <JA2>
A;Cross-references: EMBL:X75561; NID:g414687; PIDN:CAA53241.1; PID:g414689
R;James, C.M.; Gent, M.E.; Oliver, S.G.
Yeast 10, 257-264, 1994
A;Title: Sequence analysis of a 3.5 Kb EcoRI fragment from the left arm of Saccharomy
rotein.
A;Reference number: S42004; MUID:94262329
A;Accession: S42005
A;Molecule type: DNA
A;Residues: 237-1050 <JA3>
A;Cross-references: EMBL:X75561
```

```
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:BUD2; CLA2; ERC25
A;Cross-references: SGD:S0001575; MIPS:YKL092c
A;Map position: 11L
C;Superfamily: ras-specific GAP catalytic domain homology
F;520-741/Domain: ras-specific GAP catalytic domain homology <GAP>


  Query Match             13.2%;  Score 73;  DB 2;  Length 1104;
  Best Local Similarity   23.7%;  Pred. No. 18;
  Matches   23;  Conservative   17;  Mismatches   31;  Indels   26;  Gaps    3;

Qy      21 KMAETCPIFYDVFFAVANGNE---LLLDLSLTKVNATEPERTAMKKIQDCYVENGL---- 73
           |: :|  || |:| :::   |   :|:|  | |:: |   |   | :   :| | |
Db     505 KLTQTSIIFLDIFQSLSRIEEWFHVLIDKELAKIDGT-VSRINQKNLDSKHVFNSLFRGN 563

Qy      74 ------------------ISRVLDGLVMTTISSSKDC 92
                             :|: |  ::   | |:| |
Db     564 SILTKSIEQYFFRVGNEYLSKALSAILKEIIESNKSC 600


RESULT   6
T17196
IMP dehydrogenase (EC 1.1.1.205) [similarity] - Chlorobium vibrioforme
N;Alternate names: guaB1 protein; IMP:NAD+ oxidoreductase; IMPDH; inosine-5'-monophospha
C;Species: Chlorobium vibrioforme
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 16-Jun-2000
C;Accession: T17196
R;Petersen, B.L.; Moeller, M.G.; Stummann, B.M.; Henningsen, K.W.
submitted to the EMBL Data Library, January 1997
A;Description: Structure and organization of a 25 kbp region of the genome of the green
A;Reference number: Z18717
A;Accession: T17196
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-521 <PET>
A;Cross-references: EMBL:Z83933; PIDN:CAB06303.1
C;Genetics:
A;Gene: guaB
C;Function:
A;Description: catalyzes the oxidation of inosine 5'-phosphate by NAD and one molecule c
A;Pathway: purine nucleotide biosynthesis
C;Superfamily: IMP dehydrogenase; CBS homology; IMP dehydrogenase amino-terminal homolog
C;Keywords: duplication; GMP biosynthesis; NAD; oxidoreductase; purine nucleotide biosyn
F;37-103/Domain: IMP dehydrogenase amino-terminal homology <IDHN>
F;124-177/Domain: CBS homology <CBS1>
F;189-237/Domain: CBS homology <CBS2>
F;238-483/Domain: IMP dehydrogenase catalytic homology <IDHC>
F;336/Active site: Cys #status predicted


  Query Match             13.0%;  Score 72;  DB 1;  Length 521;
  Best Local Similarity   32.3%;  Pred. No. 9.6;
  Matches   21;  Conservative    9;  Mismatches   25;  Indels   10;  Gaps    2;

Qy      54 TEPERTAMKKIQDC------YVENGLISRV----LDGLVMTTISSSKDCMGEAVQNTVED 103
           :||| :: :  ||       ||  |:  |:    || :|   |   |  ||      :|:
Db     426 SEPEGSSDRYFQDASAETKKYVPEGIEGRIPQRPLDEVVYQLIGGLKSSMGYCGVKNIEE 485

Qy     104 LKLNT 108
           || ||
Db     486 LKKNT 490


RESULT   7
D86712
glycogen phosphorylase (EC 2.4.1.1) [imported] - Lactococcus lactis subsp. lactis (strai
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C;Accession: D86712
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. in press, 2001
A;Title: The complete genome sequence of the lactic acid bacterium.
A;Reference number: A86625
A;Accession: D86712
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-800 <STO>
A;Cross-references: GB:AE005176; NID:g12723610; PIDN:AAK04798.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: glgP
C;Superfamily: phosphorylase
C;Keywords: glycosyltransferase; hexosyltransferase


  Query Match             13.0%;  Score 72;  DB 2;  Length 800;
  Best Local Similarity   28.3%;  Pred. No. 16;
  Matches   17;  Conservative   14;  Mismatches   27;  Indels    2;  Gaps    1;

Qy      31 DVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTTISSSK 90
           ::| |  :|| :    |||    |  |      :| | :| :::|:|: |:  |: : |
Db     665 EIFEAAGDGNNFV--FGLTKDEVYEYYRNGNYNARDIYEQNPVVNRILNALIDGTVPNIK 722


RESULT   8
A28668
DNA mismatch repair protein mutS - Salmonella typhimurium
C;Species: Salmonella typhimurium
C;Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 29-Sep-1999
C;Accession: A28668; S61500
R;Haber, L.T.; Pang, P.P.; Sobell, D.I.; Mankovich, J.A.; Walker, G.C.
J. Bacteriol. 170, 197-202, 1988
A;Title: Nucleotide sequence of the Salmonella typhimurium mutS gene required f or mi
A;Reference number: A28668; MUID:88086868
A;Accession: A28668
A;Molecule type: DNA
A;Residues: 1-853 <HAB>
A;Cross-references: GB:M18965; NID:g154187; PIDN:AAA27167.1; PID:g154188
A;Note: the authors translated the codon CAG for residue 213 as Arg
R;Mills, D.M.; Bajaj, V.; Lee, C.A.
Mol. Microbiol. 15, 749-759, 1995
A;Title: A 40 kb chromosomal fragment encoding Salmonella typhimurium invasion genes
A;Reference number: S61499; MUID:95302986
A;Accession: S61500
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-853 <MIL>
A;Cross-references: EMBL:U16303; NID:g565091; PIDN:AAA80578.1; PID:g565092
C;Superfamily: DNA mismatch repair protein mutS
C;Keywords: DNA binding


  Query Match             12.9%;  Score 71;  DB 2;  Length 853;
  Best Local Similarity   40.4%;  Pred. No. 22;
  Matches   21;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      29 FYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           ||::|:  |     |||:|||| |:  |   |  |    ||| |   |  |
Db      38 FYELFYDDAKRASQLLDISLTKRGASAGEPIPMAGIPHHAVENYLAKLVNQG 89


RESULT   9
S77037
hypothetical protein slr0697 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S77037
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
```

```
A;Accession: S77037
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1252 <KAN>
A;Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PID:g1006579
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: hypothetical protein YKL215c


  Query Match            12.9%;  Score 71;  DB 1;  Length 1252;
  Best Local Similarity   30.4%;  Pred. No. 34;
  Matches   24;  Conservative   13;  Mismatches   38;  Indels    4;  Gaps    2;

Qy      21 KMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           |:    |  ||   || :|  ::| ||  :  |  ||: :| |:      | |: :  |
Db      49 KLLSENPELYD--DAVIHGIRILGLSSNEPIA--PEKVSMVKMGTTVATNALLEQKGDP 104

Qy      81 LVMTTISSSKDCMGEAVQN 99
           ||:      :| :  | ||
Db     105 LVLVMTQGFRDGLAIAYQN 123


RESULT  10
A72219
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: A72219
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316
A;Accession: A72219
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-295 <ARN>
A;Cross-references: GB:AE001811; GB:AE000512; NID:g4982291; PIDN:AAD36783.1; PID:g498229
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1717
C;Superfamily: conserved hypothetical protein HI1714


  Query Match            12.8%;  Score 70.5;  DB 2;  Length 295;
  Best Local Similarity   26.6%;  Pred. No. 7.1;
  Matches   25;  Conservative   19;  Mismatches   31;  Indels   19;  Gaps    4;

Qy      12 LLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVEN 71
           |:||    ||| ||     | | |:|  |||   : : |::  : :   ::|:::
Db      84 LVVT----VKMPETSTYIIDKFLVLAEKNELETVMVINKMDLYDED--DLRKVRE----- 132

Qy      72 GLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLK 105
           :  :  ||     :|:|  ||      :|:||
Db     133 --LEEIYSGLYPIVKTSAKTGMG------IEELK 158


RESULT  11
T00641
hypothetical protein F3I6.4 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Sep-2000
C;Accession: T00641
R;Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,
; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z14197
A;Accession: T00641
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-388 <FED>
A;Cross-references: EMBL:AC002396; NID:g2749918; PIDN:AAC00571.1; PID:g28865; GSPDB:GN00
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: ATSP:F3I6.4
A;Map position: 1
A;Introns: 46/2; 87/3; 115/2; 152/3; 169/3; 186/3; 223/1; 290/3
C;Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology


  Query Match            12.8%;  Score 70.5;  DB 2;  Length 388;
  Best Local Similarity   29.6%;  Pred. No. 9.8;
  Matches   24;  Conservative   12;  Mismatches   30;  Indels   15;  Gaps    3;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELL---------LDLSLTKV 51
           : : | : |  |:: ||:  |: : |  || |     :  |||         | |:|
Db     132 EALNGRVSVDPLVLGQAVSKKVEKQCAHFYAV---TISEEELLYFDQEANSGLSLALQMA 188

Qy      52 NATEPERTAMKKI---QDCYV 69
           | :||    ||:   | | |
Db     189 QAKDPETAFFKKLDGFQQCEV 209


RESULT  12
D82610
ribose-5-phosphate isomerase A XF2015 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: D82610
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: D82610
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-245 <SIM>
A;Cross-references: GB:AE004020; GB:AE003849; NID:g9107122; PIDN:AAF84817.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2015
C;Superfamily: Haemophilus influenzae ribose-5-phosphate isomerase


  Query Match            12.7%;  Score 70;  DB 2;  Length 245;
  Best Local Similarity   27.0%;  Pred. No. 6.5;
  Matches   27;  Conservative   13;  Mismatches   22;  Indels   38;  Gaps    4;

Qy      47 SLTKVNATEP--------------------ERTAMKKIQDCYVENGLISRV-------- 77
           ::|:||:|:|                    | | | |:  |||| :|  |
Db       5 TVTRVNSTQPANACAPSHIPIIINNPMSEAKRRAAEKAIE--YVENDMIIGVGTGSTVAY 62

Qy      78 -LDGLVMT------TISSSKDCMGEAVQNTVEDLKLNTLG 110
            :| |  |       :|||:       |: :| |:||  |
Db      63 FIDALGRTPKRIKGAVSSSEQSTAHLKQHGIEVLELNHTG 102


RESULT  13
I54964
DNA mismatch repair protein [imported] - Escherichia coli
```

```
C;Species: Escherichia coli
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 17-Nov-2000
C;Accession: I54964; A65054; T44995
R;Schlensog, V.; Boeck, A.
J. Bacteriol. 173, 7414-7415, 1991
A;Title: The Escherichia coli fdv gene probably encodes MutS and is located at minute 58
A;Reference number: I54964; MUID:92041649
A;Accession: I54964
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-853 <RES>
A;Cross-references: GB:M64730; NID:g146905; PIDN:AAA24188.1; PID:g146906
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: A65054
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-853 <BLAT>
A;Cross-references: GB:AE000357; GB:U00096; NID:g2367155; PIDN:AAC75775.1; PID:g1789089;
A;Experimental source: strain K-12, substrain MG1655
R;Carter, P.E.; Thomson-Carter, F.M.
submitted to the EMBL Data Library, May 1998
A;Description: Comparison of the mutS-rpoS region from Escherichia coli O157:H7, Verocyt
A;Reference number: Z22892
A;Accession: T44995
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-18,'K',20-336,'E',338-853 <CAR>
A;Cross-references: EMBL:AJ006210; PIDN:CAB43497.1
A;Experimental source: strain O157:H7
C;Genetics:
A;Gene: mutS
C;Superfamily: DNA mismatch repair protein mutS


  Query Match             12.5%;  Score 69;  DB 2;  Length 853;
  Best Local Similarity   40.4%;  Pred. No. 35;
  Matches   21;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      29 FYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           ||:::|:  |     |||:||||  |:  |   |  |    ||| |   |  |
Db      36 FYELFYDDAKRASQLLDISLTKRGASAGEPIPMAGIPYHAVENYLAKLVNQG 87


RESULT  14
F85922
methyl-directed mismatch repair [imported] - Escherichia coli (strain O157:H7)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: F85922
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85922
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-853 <STO>
A;Cross-references: GB:AE005174; NID:g12517186; PIDN:AAG57842.1; GSPDB:GN00145; UWGP:Z40
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: mutS
C;Superfamily: DNA mismatch repair protein mutS


  Query Match             12.5%;  Score 69;  DB 2;  Length 853;
  Best Local Similarity   40.4%;  Pred. No. 35;
  Matches   21;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      29 FYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           ||:::|:  |     |||:||||  |:  |   |  |    ||| |   |  |
Db      36 FYELFYDDAKRASQLLDISLTKRGASAGEPIPMAGIPYHAVENYLAKLVNQG 87


RESULT  15
T37459
ribonucleotide reductase (EC 1.17.4.-), B12-dependent - Thermoplasma acidophilum
C;Species: Thermoplasma acidophilum
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T37459
R;Tauer, A.; Benner, S.A.
Proc. Natl. Acad. Sci. U.S.A. 94, 53-58, 1997
A;Title: The B12-dependent ribonucleotide reductase from the archaebacterium Thermopl
A;Reference number: Z21699; MUID:97144393
A;Accession: T37459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-857 <TAU>
A;Cross-references: EMBL:U73619; PIDN:AAB18239.1
A;Note: cofactor vitamine B12
C;Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain
C;Keywords: oxidoreductase


  Query Match             12.5%;  Score 69;  DB 2;  Length 857;
  Best Local Similarity   27.4%;  Pred. No. 36;
  Matches   26;  Conservative   17;  Mismatches   26;  Indels   26;  Gaps    5;

Qy      30 YDVFFAVANGN----ELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTT 85
           |    :| ||     | | |  : ::   : || ::::|||  ||: |::  :||   |
Db      25 YKAMLSVKNGTMKDAEQLADKVVARLK--DKERPSVEEIQD-VVEDVLMTSKIDGKTFTD 81

Qy      86 ISSS-------------KDCMGEAVQNTVEDLKL 106
           :: |              |: ||       :||||
Db      82 VAKSYILYREKRRAIREEKELMG-----VKDDLKL 111


Search completed: May  7, 2002, 12:14:09
Job time: 116 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:13:48 ; Search time 11.68 Seconds
(without alignments)
348.441 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 552
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	541	98.0	109	1	FEL2_FELCA	P30440	felis silve
2	78	14.1	1247	1	IF3X_CAEEL	P34466	caenorhabdi
3	73	13.2	1104	1	BUD2_YEAST	P33314	saccharomyc
4	72	13.0	521	1	IMDH_CHLVI	O50316	chlorobium
5	71	12.9	854	1	MUTS_SALTY	P10339	salmonella
6	69	12.5	853	1	MUTS_ECO57	Q9s6p8	escherichia
7	69	12.5	853	1	MUTS_ECOLI	P23909	escherichia
8	68.5	12.4	375	1	MPA1_CHAOB	Q96385	chamaecypar
9	68	12.3	520	1	AMPL_ARATH	P30184	arabidopsis
10	67	12.1	587	1	BAR1_YEAST	P12630	saccharomyc
11	67	12.1	1343	1	RPOB_HAEIN	P43738	haemophilus
12	65.5	11.9	351	1	PERB_ARMRU	P15232	armoracia r
13	65.5	11.9	468	1	HEM1_MYCTU	Q11139	mycobacteri
14	65.5	11.9	1310	1	ACN1_HUMAN	O14525	homo sapien
15	65	11.8	522	1	NHAB_HAEIN	P44706	haemophilus
16	65	11.8	870	1	RPA1_METTH	O27125	methanobact
17	65	11.8	1390	1	RPOB_MYCGA	P47715	mycoplasma
18	65	11.8	2412	1	POL1_BAYMG	Q04574	barley yell
19	65	11.8	4128	1	PRKD_MOUSE	P97313	mus musculu
20	64.5	11.7	114	1	SZ05_HUMAN	P42830	homo sapien
21	64.5	11.7	936	1	ODO1_RICPR	Q9zdy3	rickettsia
22	64.5	11.7	1116	1	YK54_AQUAE	O67838	aquifex aeo
23	64	11.6	138	1	RR8_CHLVU	P56361	chlorella v
24	64	11.6	355	1	C3X1_HUMAN	P49238	homo sapien
25	64	11.6	481	1	6PGD_CERCA	P41570	ceratitis c
26	64	11.6	556	1	AGM1_ARATH	P57750	arabidopsis
27	63.5	11.5	319	1	ERA_TREPA	O83552	treponema p
28	63.5	11.5	351	1	IHBC_RAT	Q9wuk5	rattus norv
29	63.5	11.5	416	1	PGK_PLAFA	P27362	plasmodium
30	63.5	11.5	1018	1	SCA4_RICJA	Q9aj79	rickettsia
31	63	11.4	132	1	RR8_CYAPA	P14826	cyanophora
32	63	11.4	287	1	DMSC_ECOLI	P18777	escherichia
33	63	11.4	287	1	NAPH_ECOLI	P33934	escherichia
34	63	11.4	315	1	PER1_HORVU	P27337	hordeum vul
35	63	11.4	335	1	G3P_BACCO	P15115	bacillus co
36	63	11.4	350	1	DPD3_YEAST	P47110	saccharomyc
37	63	11.4	677	1	FUS2_YEAST	Q05670	saccharomyc
38	62.5	11.3	202	1	YDH1_XANAU	P22645	xanthobacte
39	62.5	11.3	237	1	SPIN_HUMAN	Q9y657	homo sapien
40	62.5	11.3	240	1	SPIN_MOUSE	Q61142	mus musculu
41	62.5	11.3	323	1	ACC1_ARATH	Q06588	arabidopsis
42	62	11.2	90	1	LPPB_HUMAN	O95969	homo sapien
43	62	11.2	134	1	Y886_HAEIN	P44069	haemophilus
44	62	11.2	517	1	GUAA_VIBCH	Q9ktw2	vibrio chol
45	62	11.2	1744	1	RPA1_TRYBB	P16355	trypanosoma

ALIGNMENTS

RESULT 1
FEL2_FELCA
ID FEL2_FELCA STANDARD; PRT; 109 AA.
AC P30440;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 2 PRECURSOR (ALLERGEN FEL D 1-B)
DE (FEL D I-B) (ALLERGEN CAT-1) (AG4) (FDI).
GN CH2.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-100.
RX MEDLINE=92052157; PubMed=1946388;
RA Morgenstern J.P., Griffith I.J., Brauer A.W., Rogers B.L.,
RA Bond J.F., Chapman M.D., Kuo M.-C.;
RT "Amino acid sequence of Fel dI, the major allergen of the domestic
RT cat: protein sequence analysis and cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:9690-9694(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92241678; PubMed=1572548;
RA Griffith I.J., Craig S., Pollock J., Yu X.-B., Morgenstern J.P.,
RA Rogers B.L.;
RT "Expression and genomic structure of the genes encoding FdI, the
RT major allergen from the domestic cat.";
RL Gene 113:263-268(1992).
RN [3]
RP SEQUENCE OF 18-37, AND CHARACTERIZATION.
RX MEDLINE=91287714; PubMed=1712068;
RA Duffort O.A., Carreira J., Nitti G., Polo F., Lombardero M.;
RT "Studies on the biochemical structure of the major cat allergen Felis
RT domesticus I.";
RL Mol. Immunol. 28:301-309(1991).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=84265679; PubMed=6747135;
RA Leiterman K., Ohman J.L. Jr.;
RT "Cat allergen 1: Biochemical, antigenic, and allergenic properties.";
RL J. Allergy Clin. Immunol. 74:147-153(1984).
CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
CC -!- ALTERNATIVE PRODUCTS: THE LONG (CH2L) AND THE SHORT (CH2S) FORMS
CC MAY ARISE BY ALTERNATIVE SPLICING OR MAY REPRESENT DIFFERENT
CC ALLELES OF THE CH2 GENE. THE SEQUENCE SHOWN HERE IS THAT OF THE
CC LONG FORM (CH2L).
CC -!- TISSUE SPECIFICITY: THE LONG FORM IS PREFERENTIALLY EXPRESSED IN
CC THE SALIVARY GLAND, WHILE THE SHORT FORM IS PREFERENTIALLY
CC EXPRESSED IN THE SKIN.
CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
CC ----------

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M77341; AAC41616.1; -.
DR   EMBL; X62478; CAA44345.1; -.
DR   PIR; JC1127; JC1127.
DR   PIR; JC1145; JC1145.
KW   Allergen; Glycoprotein; Signal; Polymorphism; Alternative splicing.
FT   SIGNAL       1    17
FT   CHAIN       18   109       MAJOR ALLERGEN I POLYPEPTIDE CHAIN 2.
FT   CARBOHYD    50    50       N-LINKED (GLCNAC. . .).
FT   VARSPLIC    82    82       T -> PSTNIAWVKQFRTP (IN ISOFORM CH2ST).
FT   VARSPLIC    83   109       MISSING (IN ISOFORM CH2ST).
FT   VARIANT     72    72       I -> L (IN CH2LV).
FT   VARIANT     72    72       I -> V (IN CH2SV).
FT   VARIANT     74    75       RV -> KF (IN CH2SV).
FT   VARIANT     82    83       TT -> IA (IN CH2S AND CH2SV).
FT   VARIANT     85    85       S -> N (IN CH2S AND CH2SV).
FT   VARIANT     86    86       S -> E (IN CH2S).
FT   VARIANT     86    86       S -> V (IN CH2SV).
FT   VARIANT     87    88       MISSING (IN CH2S AND CH2SV).
FT   VARIANT     89    89       D -> Y (IN CH2S).
FT   VARIANT     96    96       Q -> E (IN CH2SV).
FT   VARIANT    105   105       N -> K (IN CH2SV).
FT   CONFLICT    24    24       C -> F (IN REF. 3).
FT   CONFLICT    32    32       F -> T (IN REF. 3).
SQ   SEQUENCE   109 AA;  11854 MW;  857FB9CD76036CB9 CRC64;


  Query Match             98.0%;  Score 541;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-50;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109


RESULT   2
IF3X_CAEEL
ID   IF3X_CAEEL     STANDARD;      PRT;  1247 AA.
AC   P34466;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PUTATIVE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT (EIF-3).
GN   F55H2.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC       METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
CC   -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EIF-3 P135 FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z27080; CAA81605.1; -.
DR   EMBL; Z22176; CAA81605.1; JOINED.
DR   EMBL; Z22176; CAA80143.1; -.
DR   EMBL; Z27080; CAA80143.1; JOINED.
DR   WormPep; F55H2.6; CE00213.
KW   Hypothetical protein; Initiation factor; Protein biosynthesis.
SQ   SEQUENCE   1247 AA;  139920 MW;  CF44F3EF27D8774F CRC64;


  Query Match             14.1%;  Score 78;  DB 1;  Length 1247;
  Best Local Similarity   28.0%;  Pred. No. 2.9;
  Matches   26;  Conservative   14;  Mismatches   37;  Indels   16;  Gaps    2;

Qy      28 IFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLI----------SRV 77
           |  |   | | | : :|| ::  :|  | ::| |       |: | :          :
Db     384 IHADYVNAAARGVQSILDGNILAINPGEDKKTHM------YIWNNIFFSLGFDVRDHYKE 437

Qy      78 LDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLG 110
           | |      ::| |  |     |::| ||||||
Db     438 LGGDAAAFAATSTDLQGVRAFATLDDPKLNTLG 470


RESULT   3
BUD2_YEAST
ID   BUD2_YEAST     STANDARD;      PRT;  1104 AA.
AC   P33314;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   INHIBITORY REGULATOR PROTEIN BUD2/CLA2.
GN   BUD2 OR CLA2 OR ERC25 OR YKL092C OR YKL424.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94085403; PubMed=8262070;
RA   Cvrckova F., Nasmyth K.;
RT   "Yeast G1 cyclins CLN1 and CLN2 and a GAP-like protein have a role in
RT   bud formation.";
RL   EMBO J. 12:5277-5286(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93382538; PubMed=8371782;
RA   Park H.-O., Chant J., Herskowitz I.;
RT   "BUD2 encodes a GTPase-activating protein for Bud1/Rsr1 necessary for
RT   proper bud-site selection in yeast.";
RL   Nature 365:269-274(1993).
RN   [3]
RP   SEQUENCE OF 1-681 FROM N.A.
RA   Pohl T.M., Pohl F.M.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 237-1104 FROM N.A.
```

```
RX   MEDLINE=94262329; PubMed=8203166;
RA   James C.M., Gent M.E., Oliver S.G.;
RT   "Sequence analysis of a 3.5 Kb EcoRI fragment from the left arm of
RT   Saccharomyces cerevisiae chromosome XI reveals the location of the
RT   MBR1 gene and a sequence related to a GTPase-activating protein.";
RL   Yeast 10:257-264(1994).
CC   -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF BUD1/RSR1.
CC       PARTICIPATES IN THE REGULATION OF BUD-SITE SELECTION.
CC   -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 RAS-GAP DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74130; CAA52228.1; -.
DR   EMBL; L19162; AAA34461.1; -.
DR   EMBL; X75561; CAA53241.1; -.
DR   EMBL; Z28092; CAA81930.1; -.
DR   PIR; S36773; S36773.
DR   SGD; S0001575; BUD2.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001936; RasGAP.
DR   Pfam; PF00616; RasGAP; 1.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00323; RasGAP; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; FALSE_NEG.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
DR   PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR   PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW   GTPase activation.
FT   DOMAIN      322    425       C2 DOMAIN.
FT   DOMAIN      505    721       RAS-GAP.
FT   DOMAIN      496    499       POLY-SER.
FT   DOMAIN     1096   1099       POLY-LYS.
FT   CONFLICT    437    437       N -> Y (IN REF. 1).
SQ   SEQUENCE   1104 AA;  126662 MW;  451AFC3A78384760 CRC64;


  Query Match            13.2%;  Score 73;  DB 1;  Length 1104;
  Best Local Similarity  23.7%;  Pred. No. 8.4;
  Matches   23;  Conservative   17;  Mismatches   31;  Indels   26;  Gaps    3;

Qy      21 KMAETCPIFYDVFFAVANGNE---LLLDLSLTKVNATEPERTAMKKIQDCYVENGL---- 73
           |: :|  || |:| :::   |   :|:|  | |:: |   |   | :   :| | |
Db     505 KLTQTSIIFLDIFQSLSRIEEWFHVLIDKELAKIDGT-VSRINQKNLDSKHVFNSLFRGN 563

Qy      74 -----------------ISRVLDGLVMTTISSSKDC 92
                             :|: |  ::   | |:| |
Db     564 SILTKSIEQYFFRVGNEYLSKALSAILKEIIESNKSC 600


RESULT   4
IMDH_CHLVI
ID   IMDH_CHLVI     STANDARD;      PRT;   521 AA.
AC   O50316;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE   DEHYDROGENASE) (IMPDH) (IMPD).
GN   GUAB.
OS   Chlorobium vibrioforme.
OC   Bacteria; Green sulfur bacteria; Chlorobium.
OX   NCBI_TaxID=1098;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F. THIOSULFATOPHILUM NCIB 8327;
RA   Petersen B.L., Moeller M.G., Stummann B.M., Henningsen K.W.;
RT   "Clustering of genes with function in the biosynthesis of
RT   bacteriochlorophyll and heme in the green sulfur bacterium Chlorobium
RT   vibrioforme.";
RL   Hereditas 125:93-96(1996).
CC   -!- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC       XANTHOSINE 5'-PHOSPHATE + NADH.
CC   -!- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC       GMP REDUCTASE.
CC   -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z83933; CAB06303.1; -.
DR   InterPro; IPR000644; CBS.
DR   InterPro; IPR003009; FMN_enzyme.
DR   InterPro; IPR001093; IMP_DH_GMP_RED.
DR   Pfam; PF00571; CBS; 2.
DR   Pfam; PF00478; IMPDH_C; 1.
DR   Pfam; PF01574; IMPDH_N; 1.
DR   SMART; SM00116; CBS; 2.
DR   PROSITE; PS00487; IMP_DH_GMP_RED; 1.
KW   Oxidoreductase; NAD; GMP biosynthesis; Purine biosynthesis; Repeat;
KW   CBS domain.
FT   DOMAIN      118    176       CBS 1.
FT   DOMAIN      184    237       CBS 2.
FT   BINDING     336    336       IMP (POTENTIAL).
SQ   SEQUENCE   521 AA;  56627 MW;  0065FCCBC182815A CRC64;


  Query Match            13.0%;  Score 72;  DB 1;  Length 521;
  Best Local Similarity  32.3%;  Pred. No. 4.5;
  Matches   21;  Conservative    9;  Mismatches   25;  Indels   10;  Gaps    2;

Qy      54 TEPERTAMKKIQDC------YVENGLISRV----LDGLVMTTISSSKDCMGEAVQNTVED 103
           :||| :: :  ||       ||  |:  |:    || :|   |   |  ||      :|:
Db     426 SEPEGSSDRYFQDASAETKKYVPEGIEGRIPQRPLDEVVYQLIGGLKSSMGYCGVKNIEE 485

Qy     104 LKLNT 108
           || ||
Db     486 LKKNT 490


RESULT   5
MUTS_SALTY
ID   MUTS_SALTY     STANDARD;      PRT;   854 AA.
AC   P10339;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA MISMATCH REPAIR PROTEIN MUTS.
GN   MUTS.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.
RX   MEDLINE=88086868; PubMed=3275609;
RA   Haber L.T., Pang P.P., Sobell D.I., Mankovich J.A., Walker G.C.;
RT   "Nucleotide sequence of the Salmonella typhimurium mutS gene required
RT   for mismatch repair: homology of MutS and HexA of Streptococcus
RT   pneumoniae.";
RL   J. Bacteriol. 170:197-202(1988).
RN   [2]
RP   MUTAGENESIS, AND ATP-BINDING.
```

```
RX   MEDLINE=91330898; PubMed=1651234;
RA   Haber L.T., Walker G.C.;
RT   "Altering the conserved nucleotide binding motif in the Salmonella
RT   typhimurium MutS mismatch repair protein affects both its ATPase and
RT   mismatch binding activities.";
RL   EMBO J. 10:2707-2715(1991).
RN   [3]
RP   IDENTIFICATION OF PROBABLE FRAMESHIFT IN 393-424.
RX   MEDLINE=94087727; PubMed=7903399;
RA   Claverie J.-M.;
RT   "Detecting frame shifts by amino acid sequence comparison.";
RL   J. Mol. Biol. 234:1140-1157(1993).
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES
CC       IN DNA. IT IS POSSIBLE THAT IT CARRY OUT THE MISMATCH RECOGNITION
CC       STEP. THIS PROTEIN HAS A WEAK ATPASE ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U16303; AAA80578.1; -.
DR   EMBL; M18965; AAA27167.1; -.
DR   PIR; A28668; A28668.
DR   StyGene; SG10237; mutS.
DR   InterPro; IPR000432; MutS_C.
DR   InterPro; IPR002863; MutS_N.
DR   Pfam; PF00488; MutS_C; 1.
DR   Pfam; PF01624; MutS_N; 1.
DR   ProDom; PD001263; MutS_C; 1.
DR   SMART; SM00534; MUTSac; 1.
DR   SMART; SM00533; MUTSd; 1.
DR   PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW   DNA repair; ATP-binding; DNA-binding.
FT   NP_BIND     615    622       ATP (POTENTIAL).
FT   MUTAGEN     621    621       K->A: DEFECTIVE IN MISMATCH REPAIR;
FT                                ATPASE ACTIVITY REDUCED 6 FOLD.
FT   CONFLICT    393    424       ALRKKMGDFAELRDLLERAIIDAPPVLVRDGA ->
FT                                RCVKKWAISPSCATSWNAPLLTRRRYWSRRR (IN
FT                                REF. 1).
SQ   SEQUENCE   854 AA;  94757 MW;  322207EF90BBA36B CRC64;


  Query Match             12.9%;  Score 71;  DB 1;  Length 854;
  Best Local Similarity   40.4%;  Pred. No. 10;
  Matches   21;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      29 FYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           ||::|:  |     |||:|||| ||:  |   |  |    ||| |   |  |
Db      38 FYELFYDDAKRASQLLDISLTKRGASAGEPIPMAGIPHHAVENYLAKLVNQG 89


RESULT   6
MUTS_ECO57
ID   MUTS_ECO57     STANDARD;      PRT;   853 AA.
AC   Q9S6P8;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA MISMATCH REPAIR PROTEIN MUTS.
GN   MUTS OR FDV OR Z4043 OR ECS3589.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7;
RA   Leclerc J.E., Li B., Payne W.L., Cebula T.A.;
RT   "High mutation frequencies among Escherichia coli and Salmonella
RT   pathogens.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7;
RA   Carter P.E., Thomson-Carter F.M.;
RT   "Comparison of the mutS-rpoS region from Escherichia coli O157:H7,
RT   Verocytotoxin-containing and non-verocytotoxin-containing E.coli.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES
CC       IN DNA. IT IS POSSIBLE THAT IT CARRY OUT THE MISMATCH RECOGNITION
CC       STEP. THIS PROTEIN HAS A WEAK ATPASE ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U69873; AAB97931.1; -.
DR   EMBL; AJ006210; CAB43497.1; -.
DR   EMBL; AE005501; AAG57842.1; -.
DR   EMBL; AP002562; BAB37012.1; -.
DR   InterPro; IPR000432; MutS_C.
DR   InterPro; IPR002863; MutS_N.
DR   Pfam; PF00488; MutS_C; 1.
DR   Pfam; PF01624; MutS_N; 1.
DR   ProDom; PD001263; MutS_C; 1.
DR   SMART; SM00534; MUTSac; 1.
DR   SMART; SM00533; MUTSd; 1.
DR   PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW   DNA repair; ATP-binding; DNA-binding.
FT   NP_BIND     614    621       ATP (POTENTIAL).
FT   CONFLICT     59     59       G -> S (IN REF. 1).
FT   CONFLICT    144    144       G -> A (IN REF. 1).
FT   CONFLICT    317    317       D -> H (IN REF. 1).
SQ   SEQUENCE   853 AA;  95290 MW;  9D6BB82075F9B7AB CRC64;


  Query Match             12.5%;  Score 69;  DB 1;  Length 853;
  Best Local Similarity   40.4%;  Pred. No. 17;
  Matches   21;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      29 FYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           ||::|:  |     |||:|||| ||:  |   |  |    ||| |   |  |
Db      36 FYELFYDDAKRASQLLDISLTKRGASAGEPIPMAGIPYHAVENYLAKLVNQG 87
```

```
RESULT   7
MUTS_ECOLI
ID   MUTS_ECOLI     STANDARD;      PRT;   853 AA.
AC   P23909; P71279;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA MISMATCH REPAIR PROTEIN MUTS.
GN   MUTS OR FDV OR B2733.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92041649; PubMed=1938937;
RA   Schlensog V., Boeck A.;
RT   "The Escherichia coli fdv gene probably encodes mutS and is located
RT   at minute 58.8 adjacent to the hyc-hyp gene cluster.";
RL   J. Bacteriol. 173:7414-7415(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE OF 602-731 FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RA   Denamur E.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES
CC       IN DNA. IT IS POSSIBLE THAT IT CARRY OUT THE MISMATCH RECOGNITION
CC       STEP. THIS PROTEIN HAS A WEAK ATPASE ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64730; AAA24188.1; -.
DR   EMBL; U29579; AAA69243.1; -.
DR   EMBL; AE000357; AAC75775.1; -.
DR   EMBL; AF004287; AAD01197.1; -.
DR   EMBL; AF001987; AAD00921.1; -.
DR   EMBL; AF001988; AAD00922.1; -.
DR   EMBL; AF001989; AAD00923.1; -.
DR   EMBL; AF001990; AAD00924.1; -.
DR   EMBL; AF001991; AAD00925.1; -.
DR   EMBL; AF001992; AAD00926.1; -.
DR   EMBL; AF001993; AAD00927.1; -.
DR   EMBL; AF001994; AAD00928.1; -.
DR   EMBL; AF001995; AAD00929.1; -.
DR   EMBL; AF001996; AAD00930.1; -.
DR   EMBL; AF001997; AAD00931.1; -.
DR   EMBL; AF001998; AAD00932.1; -.
DR   EMBL; AF001999; AAD00933.1; -.
DR   EMBL; AF002000; AAD00934.1; -.
DR   EMBL; AF002001; AAD00935.1; -.
DR   EMBL; AF002002; AAD00936.1; -.
DR   EMBL; AF002003; AAD00937.1; -.
DR   EMBL; AF002004; AAD00938.1; -.
DR   EMBL; AF002005; AAD00939.1; -.
DR   EMBL; AF002006; AAD00940.1; -.
DR   EMBL; AF002007; AAD00941.1; -.
DR   EMBL; AF002008; AAD00942.1; -.
DR   EMBL; AF002009; AAD00943.1; -.
DR   EMBL; AF002010; AAD00944.1; -.
DR   SWISS-2DPAGE; P23909; COLI.
DR   ECO2DBASE; E093.0; 6TH EDITION.
DR   EcoGene; EG10625; mutS.
DR   InterPro; IPR000432; MutS_C.
DR   InterPro; IPR002863; MutS_N.
DR   Pfam; PF00488; MutS_C; 1.
DR   Pfam; PF01624; MutS_N; 1.
DR   ProDom; PD001263; MutS_C; 1.
DR   SMART; SM00534; MUTSac; 1.
DR   SMART; SM00533; MUTSd; 1.
DR   PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW   DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT   NP_BIND     614    621       ATP (POTENTIAL).
SQ   SEQUENCE   853 AA;  95246 MW;  5A1F295362671D55 CRC64;


  Query Match             12.5%;  Score 69;  DB 1;  Length 853;
  Best Local Similarity   40.4%;  Pred. No. 17;
  Matches   21;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      29 FYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           ||::|:  |     |||:|||| |:  |   |  |    ||| |   |  |
Db      36 FYELFYDDAKRASQLLDISLTKRGASAGEPIPMAGIPYHAVENYLAKLVNQG 87


RESULT   8
MPA1_CHAOB
ID   MPA1_CHAOB     STANDARD;      PRT;   375 AA.
AC   Q96385;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MAJOR POLLEN ALLERGEN CHA O 1 PRECURSOR.
OS   Chamaecyparis obtusa (Japanese cypress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Cupressaceae;
OC   Chamaecyparis.
OX   NCBI_TaxID=13415;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Pollen;
RX   MEDLINE=96265194; PubMed=8676896;
RA   Suzuki M., Komiyama N., Itoh M., Itoh H., Sone T., Kuno K., Takagi I.,
RA   Ohta N.;
RT   "Purification, characterization and molecular cloning of Cha o 1, a
RT   major allergen of Chamaecyparis obtusa (Japanese cypress) pollen.";
RL   Mol. Immunol. 33:451-460(1996).
CC   -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC       AMB A I/AMB A II/CRY J I SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D45404; BAA08246.1; -.
DR   Mendel; 7626; Chaob;1088;7626.
DR   InterPro; IPR002022; Amb_allergen.
DR   Pfam; PF00544; pec_lyase; 1.
DR   PRINTS; PR00807; AMBALLERGEN.
KW   Allergen; Glycoprotein; Signal.
FT   SIGNAL       1     21
FT   CHAIN       22    375       MAJOR POLLEN ALLERGEN CHA O 1.
FT   CARBOHYD    110   110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148   148       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   375 AA;  40258 MW;  81CD91DF7066DBBF CRC64;


  Query Match             12.4%;  Score 68.5;  DB 1;  Length 375;
  Best Local Similarity   27.0%;  Pred. No. 7.3;
  Matches   17;  Conservative   18;  Mismatches   23;  Indels    5;  Gaps    2;

Qy      38 NGNELLLDLS-LTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEA 96
           :|| |: : | :  |:| : :  |: : | :::  :|   ||||  |::|:    |
Db     152 SGNVLISEASGVVPVHAQDGDAITMRNVTDVWIDHNSLSDSSDGLVDVTLAST----GVT 207

Qy      97 VQN 99
           : |
Db     208 ISN 210


RESULT   9
AMPL_ARATH
ID   AMPL_ARATH     STANDARD;      PRT;   520 AA.
AC   P30184;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CYTOSOL AMINOPEPTIDASE (EC 3.4.11.1) (LEUCINE AMINOPEPTIDASE) (LAP)
DE   (LEUCYL AMINOPEPTIDASE) (PROLINE AMINOPEPTIDASE) (EC 3.4.11.5) (PROLYL
DE   AMINOPEPTIDASE).
GN   AT2G24200 OR F27D4.11.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA; TISSUE=Leaf;
RX   MEDLINE=92209533; PubMed=1555602;
RA   Bartling D., Weiler E.W.;
RT   "Leucine aminopeptidase from Arabidopsis thaliana. Molecular evidence
RT   for a phylogenetically conserved enzyme of protein turnover in higher
RT   plants.";
RL   Eur. J. Biochem. 205:425-431(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
CC       TURNOVER OF INTRACELLULAR PROTEINS. CATALYZES THE REMOVAL OF
CC       UNSUBSTITUTED AMINO-TERMINAL AMINO ACIDS FROM VARIOUS PEPTIDES.
CC   -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL AMINO ACID, XAA-|-
CC       XBB-, IN WHICH XAA IS PREFERABLY LEU, BUT MAY BE OTHER AMINO ACIDS
CC       INCLUDING PRO ALTHOUGH NOT ARG OR LYS, AND XBB MAY BE PRO.
CC   -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOHEXAMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
CC       CYTOSOL AMINOPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63444; CAA45040.1; -.
DR   EMBL; AC005967; AAD03381.1; -.
DR   PIR; S22399; S22399.
DR   HSSP; P00727; 1BPM.
DR   MEROPS; M17.002; -.
DR   InterPro; IPR000819; Peptidase_M17.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
KW   Hydrolase; Aminopeptidase; Zinc.
FT   METAL       288    288       ZINC (2) (BY SIMILARITY).
FT   METAL       293    293       ZINC (1 AND 2) (BY SIMILARITY).
FT   METAL       313    313       ZINC (2) (BY SIMILARITY).
FT   METAL       373    373       ZINC (1) (BY SIMILARITY).
FT   METAL       375    375       ZINC (1 AND 2) (BY SIMILARITY).
FT   ACT_SITE    300    300       POTENTIAL.
FT   ACT_SITE    377    377       POTENTIAL.
SQ   SEQUENCE   520 AA;  54509 MW;  D3FA9CCCD312AA92 CRC64;


  Query Match             12.3%;  Score 68;  DB 1;  Length 520;
  Best Local Similarity   28.3%;  Pred. No. 12;
  Matches   26;  Conservative   15;  Mismatches   21;  Indels   30;  Gaps    4;

Qy      46 LSLTKVNATEPERTA----------------------MKKIQDCYVENGLISRV---L 78
             | ||: |:||| : :                      : |  :   || ::|:|   |
Db       5 LGLTQPNSTEPHKISFTAKEIDVIEWKGDILVVGVTEKDLAKDGNSKFENPILSKVDAHL 64

Qy      79 DGLVMTTISSSKDCMGEAVQNTVEDLKLNTLG 110
             || :  :|| :|  |:  |:||  |:|  ||
Db      65 SGL-LAQVSSEEDFTGKPGQSTV--LRLPGLG 93


RESULT  10
BAR1_YEAST
ID   BAR1_YEAST     STANDARD;      PRT;   587 AA.
AC   P12630;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   BARRIERPEPSIN PRECURSOR (EC 3.4.23.35) (EXTRACELLULAR "BARRIER"
DE   PROTEIN) (BAR PROTEINASE).
GN   BAR1 OR SST1 OR YIL015W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88124818; PubMed=3124102;
RA   Mackay V.L., Welch S.K., Insley M.Y., Manney T.R., Holly J.,
RA   Saari G.C., Parker M.L.;
RT   "The Saccharomyces cerevisiae BAR1 gene encodes an exported protein
RT   with homology to pepsin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:55-59(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN CALLED "BARRIER ACTIVITY" IS EXCRETED BY
CC       YEAST CELLS MATING TYPE A. IT IS PROBABLY A PROTEASE THAT CLEAVES
```

```
CC       ALPHA-FACTOR AND THUS ACTS AS AN ANTAGONIST OF THIS MATING
CC       PHEROMONE AND ESTABLISHES OPTIMAL PHEROMONE CONCENTRATION FOR
CC       CONJUGATION.
CC   -!- CATALYTIC ACTIVITY: SELECTED CLEAVAGE OF 6-LEU-|-LYS-7 BOND IN
CC       THE PHEROMONE ALPHA-MATING FACTOR.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- INDUCTION: BY ALPHA-FACTOR.
CC   -!- MISCELLANEOUS: IT IS FOUND ONLY IN A MATING TYPE CELLS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC       EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC   -!- SIMILARITY: TO OTHER EUKARYOTIC ASPARTYL PROTEASES IN THE FIRST
CC       TWO THIRDS OF THE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46881; CAA86977.1; -.
DR   EMBL; J03573; AAA34451.1; -.
DR   PIR; A34084; A34084.
DR   HSSP; P28871; 1EAG.
DR   MEROPS; A01.015; -.
DR   SGD; S0001277; BAR1.
DR   InterPro; IPR001969; Asp_protease.
DR   InterPro; IPR001461; Pepsin.
DR   Pfam; PF00026; asp; 1.
DR   PRINTS; PR00792; PEPSIN.
DR   PROSITE; PS00141; ASP_PROTEASE; 2.
KW   Pheromone response; Hydrolase; Aspartyl protease; Glycoprotein;
KW   Signal.
FT   SIGNAL       1     24
FT   CHAIN       25    587       BARRIERPEPSIN.
FT   DOMAIN      25    229       PEPSIN-LIKE 1.
FT   DOMAIN     238    396       PEPSIN-LIKE 2.
FT   ACT_SITE    63     63       BY SIMILARITY.
FT   ACT_SITE   287    287       BY SIMILARITY.
FT   DISULFID   322    358       BY SIMILARITY.
FT   CARBOHYD    84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   308    308       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   366    366       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   468    468       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   503    503       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   551    551       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   587 AA;  63729 MW;  CC21DB7FDBC83984 CRC64;
```

```
  Query Match           12.1%;  Score 67;  DB 1;  Length 587;
  Best Local Similarity  25.4%;  Pred. No. 17;
  Matches   32;  Conservative   20;  Mismatches   50;  Indels   24;  Gaps    5;

Qy       2 TMRGALLVLALLVTQALGVKMAETC----------PIFYDVFFAVANGNELLLDLSLTKV 51
           |:  |   ||: : | || :   :|          |:  |   :: |  ::: |   : |
Db     249 TIVDAPATLAMTI-QGLGAQNKSSCEHETFTTTKYPVLLDSGTSLLNAPKVIADKMASFV 307

Qy      52 NATEPERTAMKKIQDCYVENG----------LISRVLDGLVMTTISSSKDCMGEAVQNT 100
           ||:  |   :  | || |  |            ||  |  |:::  :    | | ||| |
Db     308 NASYSEEEGI-YILDCPVSVGDVEYNFDFGDLQISVPLSSLILSPETEGSYC-GFAVQPT 365

Qy     101 VEDLKL 106
             : : |
Db     366 NDSMVL 371
```

```
RESULT  11
RPOB_HAEIN
ID   RPOB_HAEIN     STANDARD;      PRT;  1343 AA.
```

```
AC   P43738;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE   BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).
GN   RPOB OR HI0515.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC       ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC       BETA' CHAIN.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32733; AAC22173.1; -.
DR   TIGR; HI0515; -.
DR   InterPro; IPR001572; RNA_pol_B.
DR   Pfam; PF00562; RNA_pol_B; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase;
KW   Complete proteome.
SQ   SEQUENCE   1343 AA;  149783 MW;  4EF99CD648686A44 CRC64;
```

```
  Query Match           12.1%;  Score 67;  DB 1;  Length 1343;
  Best Local Similarity  27.3%;  Pred. No. 45;
  Matches   35;  Conservative   20;  Mismatches   37;  Indels   36;  Gaps    8;

Qy      11 ALLVTQALGVKMAETCPIFYD-VFFAVANGNELLLDLSLTKVNATEPER----TAMKKIQ 65
           | :: :|||    |   :|:| : | :| |::||: |         |||    ||   |:
Db     206 ATIILRALGYTTEEILNLFFDKITFEIA-GDKLLMTL--------VPERLRGETASFDIE 256

Qy      66 ---DCYVENG--LISRVLDGLVMTTIS------------SSKDCM----GEAVQNTVED 103
              ||| |  : :| :  |    ||             :||| :    || :     :
Db     257 ANGKVYVERGRRITARHIKALEKDNISQVVVPSEYILGKVASKDYVDLESGEIICPANGE 316

Qy     104 LKLNTLGR 111
           : | || :
Db     317 ISLETLAK 324
```

```
RESULT  12
```

```
PERB_ARMRU
ID   PERB_ARMRU     STANDARD;      PRT;   351 AA.
AC   P15232;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PEROXIDASE C1B PRECURSOR (EC 1.11.1.7).
GN   PRXC1B OR HPRC2.
OS   Armoracia rusticana (Horseradish) (Armoracia laphatifolia).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Armoracia.
OX   NCBI_TaxID=3704;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88225087; PubMed=3371352;
RA   Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,
RA   Shinmyo A., Takano M., Yamada Y., Okada H.;
RT   "Structure of the horseradish peroxidase isozyme C genes.";
RL   Eur. J. Biochem. 173:681-687(1988).
CC   -!- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,
CC       BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD
CC       WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE
CC       DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.
CC   -!- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC   -!- COFACTOR: HEME.
CC   -!- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. PLANT PEROXIDASE
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37157; AAA33378.1; -.
DR   PIR; S00626; S00626.
DR   HSSP; P00433; 2ATJ.
DR   InterPro; IPR002016; Peroxidase.
DR   Pfam; PF00141; peroxidase; 1.
DR   PRINTS; PR00458; PEROXIDASE.
DR   PRINTS; PR00461; PLPEROXIDASE.
DR   PROSITE; PS00435; PEROXIDASE_1; 1.
DR   PROSITE; PS00436; PEROXIDASE_2; 1.
KW   Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;
KW   Signal.
FT   SIGNAL        1     28
FT   CHAIN        29    351       PEROXIDASE C1B.
FT   MOD_RES      29     29       PYRROLIDONE CARBOXYLIC ACID (BY
FT                                SIMILARITY).
FT   ACT_SITE     66     66       BY SIMILARITY.
FT   ACT_SITE     70     70       DISTAL HISTIDINE (BY SIMILARITY).
FT   ACT_SITE    198    198       PROXIMAL HISTIDINE (HEME AXIAL LIGAND).
FT   DISULFID     39    119       BY SIMILARITY.
FT   DISULFID     72     77       BY SIMILARITY.
FT   DISULFID    125    329       BY SIMILARITY.
FT   DISULFID    205    237       BY SIMILARITY.
FT   CARBOHYD     41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    214    214       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    242    242       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    283    283       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   351 AA;  38645 MW;  7A8C606A3928950B CRC64;


  Query Match           11.9%;  Score 65.5;  DB 1;  Length 351;
  Best Local Similarity  26.9%;  Pred. No. 14;
  Matches   25;  Conservative   18;  Mismatches   41;  Indels    9;  Gaps    5;

Qy       7 LLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA---MKK 63
           |: |  |   |  :  |:  | |||   :  | : :: |: : ::  ::|  ||
Db      13 LITLGCLAFYA-SLSDAQLTPTFYDT--SCPNVSNIVRDIIINELR-SDPRITASILRLH 68

Qy      64 IQDCYVENGLISRVLDGLVMTTISSSKDCMGEA 96
             ||:|     | :||    |:  : || :| |
Db      69 FHDCFVNGCDASILLDN--TTSFLTEKDALGNA 99


RESULT  13
HEM1_MYCTU
ID   HEM1_MYCTU     STANDARD;      PRT;   468 AA.
AC   Q11139;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GLUTAMYL-TRNA REDUCTASE (EC 1.2.1.-) (GLUTR).
GN   HEMA OR RV0509 OR MT0530 OR MTCY20G9.36.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: GLUTAMYL-TRNA(GLU) + NADPH = GLUTAMATE-1-
CC       SEMIALDEHYDE + NADP(+) + TRNA(GLU).
CC   -!- PATHWAY: FIRST STEP IN PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMYL-TRNA REDUCTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z77162; CAB00935.1; -.
DR   EMBL; AE006953; AAK44753.1; ALT_INIT.
DR   TIGR; MT0530; -.
DR   TubercuList; Rv0509; -.
DR   InterPro; IPR000343; GlutR.
DR   InterPro; IPR000594; ThiF_family.
DR   Pfam; PF00745; GlutR; 1.
DR   PROSITE; PS00747; GLUTR; 1.
KW   Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
SQ   SEQUENCE   468 AA;  49361 MW;  DDC257648FF6EAAB CRC64;


  Query Match           11.9%;  Score 65.5;  DB 1;  Length 468;
```

```
  Best Local Similarity   23.3%;  Pred. No. 19;
  Matches   30;  Conservative   18;  Mismatches   52;  Indels   29;  Gaps    5;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFY--------DVFFAVANGNELLLDLSLTKVN 52
           |   | | |:  :: :  |: | |     |         :  ||||:|    ||      :
Db      60 DAFHGGLSVIGQVLAEHSGMSMGELTKYAYVRYSEAAVEHLFAVASG----LD------S 109

Qy      53 ATEPERTAMKKIQDCYV---ENGLISRVLDGLVMTTISSSK--------DCMGEAVQNTV 101
           |   |:  : :::  |     |  : |||  |    :|  |        |  | :| :
Db     110 AVIGEQQVLGQVRRAYAVAESNRTVGRVLHELAQRALSVGKRVHSETAIDAAGASVVSVA 169

Qy     102 EDLKLNTLG 110
             :    ||
Db     170 LGMAERKLG 178


RESULT  14
ACN1_HUMAN
ID   ACN1_HUMAN     STANDARD;      PRT;  1310 AA.
AC   O14525; O60799;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ASTROTACTIN 1 (FRAGMENT).
GN   ASTN1 OR ASTN OR KIAA0289.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=97323006; PubMed=9179496;
RA   Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M.,
RA   Seki N., Nomura N.;
RT   "Construction and characterization of human brain cDNA libraries
RT   suitable for analysis of cDNA clones encoding relatively large
RT   proteins.";
RL   DNA Res. 4:53-59(1997).
RN   [2]
RP   SEQUENCE OF 166-906 FROM N.A.
RA   Pearce A.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NEURONAL ADHESION MOLECULE THAT IS REQUIRED FOR GLIAL-
CC       GUIDED MIGRATION OF YOUNG POSTMITOTIC NEUROBLASTS IN CORTICAL
CC       REGIONS OF DEVELOPING BRAIN, INCLUDING CEREBRUM, HIPPOCAMPUS,
CC       CEREBELLUM AND OLFACTORY BULB.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 AND 2 (SHOWN HERE); MAY BE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006627; BAA22958.1; -.
DR   EMBL; AL022145; CAA18140.1; -.
DR   EMBL; AL022145; CAA18141.1; -.
DR   MIM; 600904; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001777; FN_III.
DR   InterPro; IPR001862; MAC_perforin.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00457; MACPF; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   EGF-like domain; Repeat; Transmembrane; Alternative splicing.
FT   NON_TER       1      1
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM    162    182       POTENTIAL.
FT   TRANSMEM    392    410       POTENTIAL.
FT   DOMAIN      467    515       EGF-LIKE 1.
FT   DOMAIN      616    660       EGF-LIKE 2.
FT   DOMAIN      664    716       EGF-LIKE 3.
FT   DOMAIN      804    885       FIBRONECTIN TYPE-III 1.
FT   DOMAIN     1038   1153       FIBRONECTIN TYPE-III 2.
FT   DISULFID    471    483       BY SIMILARITY.
FT   DISULFID    479    498       BY SIMILARITY.
FT   DISULFID    500    514       BY SIMILARITY.
FT   DISULFID    620    633       BY SIMILARITY.
FT   DISULFID    627    644       BY SIMILARITY.
FT   DISULFID    646    659       BY SIMILARITY.
FT   DISULFID    668    680       BY SIMILARITY.
FT   DISULFID    676    700       BY SIMILARITY.
FT   DISULFID    702    715       BY SIMILARITY.
FT   CARBOHYD    123    123       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    461    461       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    737    737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    750    750       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    812    812       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    992    992       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    488    496       MISSING (IN ISOFORM 1).
SQ   SEQUENCE   1310 AA;  145670 MW;  4A2BA8E07E03EC66 CRC64;


  Query Match             11.9%;  Score 65.5;  DB 1;  Length 1310;
  Best Local Similarity   20.8%;  Pred. No. 63;
  Matches   26;  Conservative   20;  Mismatches   38;  Indels   41;  Gaps    4;

Qy       2 TMRGALLVLALLVTQALGVK------MAETCP---------------IFYDVFFAVANG 39
           |:   |:    :    |||       : ||||               :| ::||   |
Db     692 TLYNILMFCGCIEDYKLGVDGRSCQLITETCPEGSDCGESRELPMNQTLFGEMFFGYNNH 751

Qy      40 NELLLDLSLTKVNATEPERTAMKKIQDCYVEN----GLISRVLDGLVMTTISSSKDCMGE 95
           ::              |  | : ::  : :|    ||  :: ||||: |:     |: |
Db     752 SK--------------EVAAGQVLKGTFRQNNFARGLDQQLPDGLVVATVPLENQCLEE 796

Qy      96 AVQNT 100
             : |
Db     797 ISEPT 801


RESULT  15
NHAB_HAEIN
ID   NHAB_HAEIN     STANDARD;      PRT;   522 AA.
AC   P44706;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NA(+)/H(+) ANTIPORTER 2 (SODIUM/PROTON ANTIPORTER 2).
GN   NHAB OR HI0427.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
```

```
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: THIS IS A NA+/H+ ANTIPORTER. IT HAS A HIGH AFFINITY FOR
CC       NA+ AND A RELATIVE INDEPENDENCE OF ITS ACTIVITY ON PH
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO ENTEROCOCCUS HIRAE NA(+)/H(+) ANTIPORTER
CC       (NAPA), AND LOCAL, TO E.COLI KEFB/KEFC N-TERMINAL REGION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32726; AAC22086.1; -.
DR   TIGR; HI0427; -.
KW   Transmembrane; Inner membrane; Transport; Sodium transport; Symport;
KW   Complete proteome.
FT   TRANSMEM     29     49       POTENTIAL.
FT   TRANSMEM     51     71       POTENTIAL.
FT   TRANSMEM     96    116       POTENTIAL.
FT   TRANSMEM    151    171       POTENTIAL.
FT   TRANSMEM    211    231       POTENTIAL.
FT   TRANSMEM    247    267       POTENTIAL.
FT   TRANSMEM    312    332       POTENTIAL.
FT   TRANSMEM    333    353       POTENTIAL.
FT   TRANSMEM    357    377       POTENTIAL.
FT   TRANSMEM    398    418       POTENTIAL.
FT   TRANSMEM    456    476       POTENTIAL.
FT   TRANSMEM    492    512       POTENTIAL.
SQ   SEQUENCE   522 AA;  57533 MW;  4DFF4DCD30304B97 CRC64;


  Query Match             11.8%;  Score 65;  DB 1;  Length 522;
  Best Local Similarity   27.4%;  Pred. No. 25;
  Matches   29;  Conservative   21;  Mismatches   36;  Indels   20;  Gaps    6;

Qy       6 ALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELL--LDLSLTKVNATEPERTAMKK 63
           || |:|:::: |:|          || |:  ||:|| |:  :|:|  :    | :   ::|
Db     154 ALTVVAVIISVAMG---------FYGVYHKVASGNNLIDAVDISDDR-KIIEQQHEILEK 203

Qy      64 IQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTV--EDLKLN 107
             :       | | ::   | | :      :||  || :  |  | |
Db     204 FRAF-----LRSLMMHAGVGTALGGVMTVVGEP-QNLIIAEQAKWN 243


Search completed: May  7, 2002, 12:17:05
Job time: 197 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:13:33 ; Search time 22.39 Seconds
(without alignments)
725.155 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 552
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	154	27.9	112	11	Q9JI02	Q9ji02 mus musculu
2	78	14.1	1247	5	Q9TYW0	Q9tyw0 caenorhabdi
3	76.5	13.9	2771	5	Q26216	Q26216 plasmodium
4	75	13.6	945	10	Q9M7A8	Q9m7a8 nicotiana t
5	72	13.0	404	2	Q9A0D8	Q9a0d8 streptococc
6	72	13.0	800	2	Q9CHM8	Q9chm8 lactococcus
7	71	12.9	119	11	Q9ES84	Q9es84 mus musculu
8	71	12.9	1252	2	Q55959	Q55959 synechocyst
9	70.5	12.8	295	2	Q9X242	Q9x242 thermotoga
10	70.5	12.8	388	10	O48678	O48678 arabidopsis
11	70	12.7	245	2	Q9PBX0	Q9pbx0 xylella fas
12	69	12.5	770	1	Q9HI70	Q9hi70 thermoplasm
13	69	12.5	857	1	P74911	P74911 thermoplasm
14	69	12.5	895	1	Q9YAU1	Q9yau1 aeropyrum p
15	68.5	12.4	343	4	Q92519	Q92519 homo sapien
16	68.5	12.4	343	6	Q28283	Q28283 canis famil
17	68.5	12.4	466	1	O57878	O57878 pyrococcus
18	68.5	12.4	534	10	Q9CAB0	Q9cab0 arabidopsis
19	68.5	12.4	1167	10	O49471	O49471 arabidopsis
20	68	12.3	133	8	Q9TLU5	Q9tlu5 cyanidium c
21	67.5	12.2	290	3	Q9P827	Q9p827 candida alb
22	67.5	12.2	346	10	Q9LDA4	Q9lda4 arabidopsis
23	67.5	12.2	626	5	Q9NDD1	Q9ndd1 leishmania
24	67	12.1	530	12	Q9DWD8	Q9dwd8 rat cytomeg
25	67	12.1	675	10	Q9FNM5	Q9fnm5 arabidopsis
26	66.5	12.0	378	3	Q9USZ2	Q9usz2 schizosacch
27	66.5	12.0	488	5	Q9VM95	Q9vm95 drosophila
28	66	12.0	468	4	O00220	O00220 homo sapien
29	66	12.0	636	12	O90142	O90142 spodoptera
30	66	12.0	1063	12	Q9J844	Q9j844 spodoptera
31	65.5	11.9	313	2	Q9EWU7	Q9ewu7 streptomyce
32	65.5	11.9	320	10	Q9ZWP2	Q9zwp2 lycopersico
33	65.5	11.9	352	10	O22443	O22443 glycine max
34	65.5	11.9	5825	10	O82731	O82731 vicia faba
35	65	11.8	270	2	Q9A0H2	Q9a0h2 streptococc
36	65	11.8	326	10	Q9C8U7	Q9c8u7 arabidopsis
37	65	11.8	343	10	Q9SNV1	Q9snv1 antirrhinum
38	65	11.8	392	2	Q9K0N8	Q9k0n8 neisseria m
39	65	11.8	392	2	Q9JVR6	Q9jvr6 neisseria m
40	65	11.8	400	5	Q9W115	Q9w115 drosophila
41	65	11.8	483	5	Q9VXS8	Q9vxs8 drosophila
42	65	11.8	530	2	Q9KQU7	Q9kqu7 vibrio chol
43	65	11.8	571	5	O97446	O97446 giardia lam
44	65	11.8	574	2	Q9AA11	Q9aa11 caulobacter
45	64.5	11.7	157	6	O62841	O62841 equus cabal

ALIGNMENTS

RESULT 1
Q9JI02
ID Q9JI02 PRELIMINARY; PRT; 112 AA.
AC Q9JI02;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE 11 KDA SECRETED PROTEIN PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=EXORBITAL LACRIMAL GLAND;
RA Remington S.G., Nelson J.D.;
RT "Predicted 11kD secreted protein related to major allergen Fel dI
RT chain 2.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF272844; AAF81793.1; -.
SQ SEQUENCE 112 AA; 12589 MW; 838A68DC00433BAF CRC64;

Query Match 27.9%; Score 154; DB 11; Length 112;
Best Local Similarity 35.2%; Pred. No. 1.4e-08;
Matches 32; Conservative 23; Mismatches 36; Indels 0; Gaps 0;

```
Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |:| ||:| |||| |  :  | |  |::: : :| :|: : |   |   ||| |: |::
Db       1 MKGTLLLLGLLVTGELSFQTTEACLPFFEGYASVLSGSRVWLYQELQAFNATAEEKVALE 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCM 93
           |||||: |  : : :|:  :|  : :| :|:
Db      61 KIQDCFSEERIRNILLEPKIMEAMVASPECL 91
```

RESULT 2
Q9TYW0
ID Q9TYW0 PRELIMINARY; PRT; 1247 AA.
AC Q9TYW0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CLU-1.
GN   CLU-1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RA   Wickham J.Q., Moulder G., Barstead R., Clarke M.;
RT   "Cloning and disruption of the clu-1 gene of C. elegans.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF101078; AAC72406.1; -.
SQ   SEQUENCE   1247 AA;  139992 MW;  BC2591EF25D9FAF8 CRC64;


  Query Match             14.1%;  Score 78;  DB 5;  Length 1247;
  Best Local Similarity   28.0%;  Pred. No. 19;
  Matches   26;  Conservative   14;  Mismatches   37;  Indels   16;  Gaps    2;

Qy      28 IFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLI----------SRV 77
           |  |    | | | : :|| ::  :|  | ::| |      |: | :           :
Db     384 IHADYVNAAARGVQSILDGNILAINPGEDKKTHM------YIWNNIFFSLGFDVRDHYKE 437

Qy      78 LDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLG 110
           | |      ::| |  |     ||::| ||||||
Db     438 LGGDAAAFAATSTDLQGVRAFATLDDPKLNTLG 470


RESULT   3
Q26216
ID   Q26216      PRELIMINARY;      PRT;  2771 AA.
AC   Q26216;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   RHOPTRY PROTEIN.
OS   Plasmodium yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5861;
RN   [1]
RP   SEQUENCE OF 379-2771 FROM N.A.
RC   STRAIN=YM;
RX   MEDLINE=97077455; PubMed=8920022;
RA   Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;
RT   "Comparison of two members of a multigene family coding for high-
RT   molecular mass rhoptry proteins of Plasmodium yoelii.";
RL   Mol. Biochem. Parasitol. 76:329-332(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YM;
RA   Green J.L., Holder A.A.;
RT   "Structure of the E8 gene encoding a high molecular mass rhoptry
RT   protein of Plasmodium yoelii.";
RL   Mol. Biochem. Parasitol. 0:0-0(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YM;
RA   Holder A.A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U36927; AAB41263.3; -.
SQ   SEQUENCE   2771 AA;  325640 MW;  C0CCB9AB6E7ACF36 CRC64;


  Query Match             13.9%;  Score 76.5;  DB 5;  Length 2771;
  Best Local Similarity   35.1%;  Pred. No. 72;
  Matches   27;  Conservative   12;  Mismatches   25;  Indels   13;  Gaps    4;

Qy      34 FAVANGNELLLDLSLTKV-NATEPERTAMKKIQDCYVE-NGLISRVLDGLVMTTISSSKD 91
           |:  | | |     | :: ||   |   :|||:|  :| ||||:::        | : ||
Db    2102 FSEENNNILQSKKKLKELTNAFNAE---IKKIEDKIIEKNGLINKL--------IETRKD 2150

Qy      92 CMGEAVQNTVEDLKLNT 108
           ||     :  || ||: |
Db    2151 CMLFTYKTLVETLKIKT 2167


RESULT   4
Q9M7A8
ID   Q9M7A8      PRELIMINARY;      PRT;  945 AA.
AC   Q9M7A8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LRR RECEPTOR-LIKE PROTEIN KINASE.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. XANTHI;
RA   Cho H.S., Yoon G.M., Liu J.R., Pai Lee H.S.;
RT   "Expression of NtRLK1 encoding a LRR receptor-like kinase is
RT   stimulated by abiotic and biotic stress.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AF142596; AAF66615.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR002290; Ser_thr_kin_actsite.
DR   InterPro; IPR001245; Tyr_kin.
DR   Pfam; PF00560; LRR; 6.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   PRINTS; PR00109; TYRKINASE.
DR   SMART; SM00370; LRR; 7.
DR   SMART; SM00369; LRR_TYP; 1.
DR   SMART; SM00221; STYKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
KW   Transferase.
SQ   SEQUENCE   945 AA;  101787 MW;  30741AFE8909145E CRC64;


  Query Match             13.6%;  Score 75;  DB 10;  Length 945;
  Best Local Similarity   26.1%;  Pred. No. 28;
  Matches   24;  Conservative   18;  Mismatches   38;  Indels   12;  Gaps    4;

Qy      17 ALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISR 76
           |:   :  | | |:| | :: |     | ||   :  : |   :  :||  :: ||| |
Db     174 AVSANITGTIPDFFDAFASLTN-----LHLSFNNLGGSLPSSFSGSQIQSLWL-NGLKGR 227

Qy      77 VLDGLV-----MTTISSSKDCMGEAVQNTVED 103
           |:| :     || :: :  |   |  : : |
Db     228 -LNGSIAVIQNMTQLTRTSGCKANAFSSPLPD 258


RESULT   5
Q9A0D8
ID   Q9A0D8      PRELIMINARY;      PRT;  404 AA.
AC   Q9A0D8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE THIAMINE BIOSYNTHESIS PROTEIN.
GN   THII OR SPY0817.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
```

```
OX   NCBI_TaxID=1314;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SF370;
RX   MEDLINE=21192684; PubMed=11296296;
RA   Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA   Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA   Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
RA   Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT   "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR   EMBL; AE006532; AAK33753.1; -.
KW   Complete proteome.
SQ   SEQUENCE   404 AA;  44730 MW;  9BC8B801385463AB CRC64;


  Query Match             13.0%;  Score 72;  DB 2;  Length 404;
  Best Local Similarity   27.5%;  Pred. No. 21;
  Matches   14;  Conservative   17;  Mismatches   18;  Indels    2;  Gaps    1;

Qy      54 TEPERTAMKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDL 104
           | |:    :| ::|:  :||: | : |:|:| |  : : : : |:| :: |
Db     355 TNPKLGNAEKYEECFDIDGLVQRAVSGIVVTEI--TPEIVNDEVENLIDAL 403


RESULT   6
Q9CHM8
ID   Q9CHM8      PRELIMINARY;      PRT;   800 AA.
AC   Q9CHM8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GLYCOGEN PHOSPHORYLASE (EC 2.4.1.1) (ALPHA-GLUCAN PHOSPHORYLASE)
DE   (STARCH PHOSPHORYLASE).
GN   GLGP.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis.";
RL   Genome Res. 0:0-0(2001).
CC   -!- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN
CC       CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN
CC       THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.
CC       HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL
CC       PROPERTIES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: (1,4-ALPHA-D-GLUCOSYL)(N) + ORTHOPHOSPHATE =
CC       (1,4-ALPHA-D-GLUCOSYL)(N-1) + ALPHA-D-GLUCOSE 1-PHOSPHATE.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SIMILARITY: TO THE GLYCOGEN PHOSPHORYLASE FAMILY.
DR   EMBL; AE006303; AAK04798.1; -.
DR   InterPro; IPR000811; Phosphorylase.
DR   Pfam; PF00343; phosphorylase; 1.
DR   PROSITE; PS00102; PHOSPHORYLASE; 1.
KW   Carbohydrate metabolism; Complete proteome; Glycosyltransferase;
KW   Pyridoxal phosphate; Transferase.
SQ   SEQUENCE   800 AA;  91812 MW;  A9A3CC9A81D0AD27 CRC64;


  Query Match             13.0%;  Score 72;  DB 2;  Length 800;
  Best Local Similarity   28.3%;  Pred. No. 48;
  Matches   17;  Conservative   14;  Mismatches   27;  Indels    2;  Gaps    1;

Qy      31 DVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTTISSSK 90
           ::| |  :||  :    |||    |  |      :| | :| :::|:|: |:  |: : |
Db     665 EIFEAAGDGNNFV--FGLTKDEVYEYYRNGNYNARDIYEQNPVVNRILNALIDGTVPNIK 722


RESULT   7
Q9ES84
ID   Q9ES84      PRELIMINARY;      PRT;   119 AA.
AC   Q9ES84;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SNRNP CORE PROTEIN SMX5D (FRAGMENT).
GN   SMX5D.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.D2/SGSNJ; TISSUE=TESTIS;
RA   Zhang J., Ma Z.R., Meeker N.D., Teuscher C.;
RT   "Identification and characterization of alternatively spliced SmX5
RT   isoforms in mouse.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF204157; AAG31434.1; -.
DR   InterPro; IPR001163; snRNP_Sm.
DR   Pfam; PF01423; Sm; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   119 AA;  13228 MW;  B326E5EC16D6F204 CRC64;


  Query Match             12.9%;  Score 71;  DB 11;  Length 119;
  Best Local Similarity   23.2%;  Pred. No. 6.2;
  Matches   22;  Conservative   21;  Mismatches   28;  Indels   24;  Gaps    4;

Qy       6 ALLVLALLVTQALGVKMAETCP---IFYDVFFAVANGNELL------------------ 43
             ||: :|   :|| |  :   ||   :||  ||    | :::
Db       1 ALVSVAWGGSQAPGFSVQPYCPHPSLFYS-FFKSLVGKDVVVELKNDLSICGTLHSVDQY 59

Qy      44 LDLSLTKVNATEPER-TAMKKIQDCYVENGLISRV 77
           |:: || :: |:||:   |  :::|::   ::  |
Db      60 LNIKLTDISVTDPEKYPHMLSVKNCFIRGSVVRYV 94


RESULT   8
Q55959
ID   Q55959      PRELIMINARY;      PRT;  1252 AA.
AC   Q55959;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 136.1 KDA PROTEIN.
GN   SLR0697.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
```

```
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D64005; BAA10729.1; -.
DR   InterPro; IPR002821; Hydantoinase_A.
DR   InterPro; IPR003692; Hydantoinase_B.
DR   Pfam; PF01968; Hydantoinase_A; 1.
DR   Pfam; PF02538; Hydantoinase_B; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   1252 AA;  136051 MW;  07ECCA495CD3E25F CRC64;


  Query Match             12.9%;  Score 71;  DB 2;  Length 1252;
  Best Local Similarity   30.4%;  Pred. No. 1e+02;
  Matches   24;  Conservative   13;  Mismatches   38;  Indels    4;  Gaps    2;

Qy      21 KMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDG 80
           |:    |  ||   || :|  ::|| ||  :  |  ||: :| |:      | |: :  |
Db      49 KLLSENPELYD--DAVIHGIRIILGLSSNEPIA--PEKVSMVKMGTTVATNALLEQKGDP 104

Qy      81 LVMTTISSSKDCMGEAVQN 99
           ||:      :| :  | ||
Db     105 LVLVMTQGFRDGLAIAYQN 123


RESULT   9
Q9X242
ID   Q9X242      PRELIMINARY;      PRT;   295 AA.
AC   Q9X242;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   TM1717.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001811; AAD36783.1; -.
DR   TIGR; TM1717; -.
KW   Complete proteome.
SQ   SEQUENCE   295 AA;  33641 MW;  E8A9E7396C855848 CRC64;


  Query Match             12.8%;  Score 70.5;  DB 2;  Length 295;
  Best Local Similarity   26.6%;  Pred. No. 21;
  Matches   25;  Conservative   19;  Mismatches   31;  Indels   19;  Gaps    4;

Qy      12 LLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVEN 71
           |:||    ||| ||     | | |:|  |||   : : |::  : :   ::|:::
Db      84 LVVT----VKMPETSTYIIDKFLVLAEKNELETVMVINKMDLYDED--DLRKVRE----- 132

Qy      72 GLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLK 105
             :  :  ||     :|:|  ||      :|:||
Db     133 --LEEIYSGLYPIVKTSAKTGMG------IEELK 158


RESULT  10
O48678
ID   O48678      PRELIMINARY;      PRT;   388 AA.
AC   O48678;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F3I6.4 PROTEIN.
GN   F3I6.4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA   Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
RA   Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,
RA   Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002396; AAC00573.1; -.
DR   HSSP; P25685; 1HDJ.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS00636; DNAJ_1; UNKNOWN_1.
DR   PROSITE; PS50076; DNAJ_2; 1.
SQ   SEQUENCE   388 AA;  44023 MW;  DD815C3AD9B0AF01 CRC64;


  Query Match             12.8%;  Score 70.5;  DB 10;  Length 388;
  Best Local Similarity   29.6%;  Pred. No. 29;
  Matches   24;  Conservative   12;  Mismatches   30;  Indels   15;  Gaps    3;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELL---------LDLSLTKV 51
           : : | : |  |:: ||:  |: : |  || |     :  |||         | |:|
Db     132 EALNGRVSVDPLVLGQAVSKKVEKQCAHFYAV---TISEEELLYFDQEANSGLSLALQMA 188

Qy      52 NATEPERTAMKKI---QDCYV 69
              | :||    ||:   | | |
Db     189 QAKDPETAFFKKLDGFQQCEV 209


RESULT  11
Q9PBX0
ID   Q9PBX0      PRELIMINARY;      PRT;   245 AA.
AC   Q9PBX0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   RIBOSE-5-PHOSPHATE ISOMERASE A.
GN   XF2015.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
```

```
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004020; AAF84817.1; -.
KW   Complete proteome.
SQ   SEQUENCE   245 AA;  26561 MW;  2D2079AD223900E5 CRC64;


  Query Match             12.7%;  Score 70;  DB 2;  Length 245;
  Best Local Similarity   27.0%;  Pred. No. 19;
  Matches   27;  Conservative   13;  Mismatches   22;  Indels   38;  Gaps    4;

Qy      47 SLTKVNATEP--------------------ERTAMKKIQDCYVENGLISRV-------- 77
           ::|:||:|:|                    | | | |:  |||| :|  |
Db       5 TVTRVNSTQPANACAPSHIPIIINNPMSEAKRRAAEKAIE--YVENDMIIGVGTGSTVAY 62

Qy      78 -LDGLVMT------TISSSKDCMGEAVQNTVEDLKLNTLG 110
            :| |  |       :|||:       |: :| |:||  |
Db      63 FIDALGRTPKRIKGAVSSSEQSTAHLKQHGIEVLELNHTG 102


RESULT  12
Q9HI70
ID   Q9HI70      PRELIMINARY;      PRT;   770 AA.
AC   Q9HI70;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   RIBONUCLEOTIDE REDUCTASE.
GN   TA1475.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445067; CAC12593.1; -.
DR   InterPro; IPR000788; Ribonucleo_red.
DR   Pfam; PF00317; ribonucleo_red; 3.
DR   PRINTS; PR01183; RIBORDTASEM1.
KW   Complete proteome.
SQ   SEQUENCE   770 AA;  86925 MW;  4A5E1E1945842CF4 CRC64;


  Query Match             12.5%;  Score 69;  DB 1;  Length 770;
  Best Local Similarity   27.4%;  Pred. No. 93;
  Matches   26;  Conservative   17;  Mismatches   26;  Indels   26;  Gaps    5;

Qy      30 YDVFFAVANGN----ELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTT 85
           |    :| ||     | | |  : ::   : || ::::||| ||: |::  :||   |
Db      27 YKAMLSVKNGTMKDAEQLADKVVARLK--DKERPSVEEIQD-VVEDVLMTSKIDGKTFTD 83

Qy      86 ISSS--------------KDCMGEAVQNTVEDLKL 106
           :: |              |: ||       :||||
Db      84 VAKSYILYREKRRAIREEKELMG-----VKDDLKL 113


RESULT  13
P74911
ID   P74911      PRELIMINARY;      PRT;   857 AA.
AC   P74911;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   RIBONUCLEOTIDE REDUCTASE.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tauer A., Benner S.A.;
RL   Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
DR   EMBL; U73619; AAB18239.1; -.
DR   InterPro; IPR000788; Ribonucleo_red.
DR   Pfam; PF00317; ribonucleo_red; 3.
DR   PRINTS; PR01183; RIBORDTASEM1.
SQ   SEQUENCE   857 AA;  96960 MW;  9144086F575B2B93 CRC64;


  Query Match             12.5%;  Score 69;  DB 1;  Length 857;
  Best Local Similarity   27.4%;  Pred. No. 1.1e+02;
  Matches   26;  Conservative   17;  Mismatches   26;  Indels   26;  Gaps    5;

Qy      30 YDVFFAVANGN----ELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTT 85
           |    :| ||     | | |  : ::   : || ::::||| ||: |::  :||   |
Db      25 YKAMLSVKNGTMKDAEQLADKVVARLK--DKERPSVEEIQD-VVEDVLMTSKIDGKTFTD 81

Qy      86 ISSS--------------KDCMGEAVQNTVEDLKL 106
           :: |              |: ||       :||||
Db      82 VAKSYILYREKRRAIREEKELMG-----VKDDLKL 111


RESULT  14
Q9YAU1
ID   Q9YAU1      PRELIMINARY;      PRT;   895 AA.
AC   Q9YAU1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   895AA LONG HYPOTHETICAL DNA-DIRECTED RNA POLYMERASE SUBUNIT A'.
GN   APE1853.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000062; BAA80857.1; -.
DR   InterPro; IPR000722; RNA_pol_A.
DR   Pfam; PF00623; RNA_pol_A; 1.
KW   DNA-directed RNA polymerase; Complete proteome.
SQ   SEQUENCE   895 AA;  101740 MW;  1E01BC06B4BD0AAE CRC64;


  Query Match             12.5%;  Score 69;  DB 1;  Length 895;
  Best Local Similarity   30.9%;  Pred. No. 1.1e+02;
  Matches   21;  Conservative   12;  Mismatches   21;  Indels   14;  Gaps    2;

Qy      31 DVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLDGLVMTTISSSK 90
```

```
          :|||  | | | |:|          ||:|  |  |::  ||: : | :|:|   : :
Db    806 EVFFHAAAGREGLVD----------TAVKTSQSGYMQRRLINALQD-IVVTYDGTVR 851

Qy     91 DCMGEAVQ 98
          |  |  :|
Db    852 DLYGNLIQ 859


RESULT  15
Q92519
ID   Q92519      PRELIMINARY;      PRT;   343 AA.
AC   Q92519;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GS3955 (GS3955 PROTEIN).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CANCELLOUS BONE;
RA   Ohno I., Hashimoto J., Takaoka K., Ochi T., Okubo K., Matsubara K.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ENDOMETRIAL ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D87119; BAA13250.1; -.
DR   EMBL; BC002637; AAH02637.1; -.
DR   HSSP; P08631; 1AD5.
DR   InterPro; IPR000719; Euk_pkinase.
DR   Pfam; PF00069; pkinase; 2.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   343 AA;  38800 MW;  BF8B7366DACB84FA CRC64;


  Query Match             12.4%;  Score 68.5;  DB 4;  Length 343;
  Best Local Similarity   35.2%;  Pred. No. 40;
  Matches   19;  Conservative    9;  Mismatches   21;  Indels    5;  Gaps    2;

Qy     24 ETCPIFYDVFFAVA---NGNELLLDLSLTKVNATEPERTAMK--KIQDCYVENG 72
          |   :|| :  |||   :|  :|||| | |    : ||| :|   ::| |:  |
Db    151 EAARLFYQIASAVAHCHDGGLVLRDLKLRKFIFKDEERTRVKLESLEDAYILRG 204


Search completed: May  7, 2002, 12:16:47
Job time: 194 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:14:13 ; Search time 23.78 Seconds
(without alignments)
345.758 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 111
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	111	100.0	111	13	AAR27369	TRFP Chain #2 with
2	111	100.0	111	21	AAB28933	T cell reactive fe
3	111	100.0	111	21	AAY87673	Feline human TRFP
4	111	100.0	111	21	AAY90103	Cat TRFP chain 2 l
5	111	100.0	111	21	AAY51470	Human TRFP chain 2
6	109	98.2	109	14	AAR36541	TRFP chain 2 (with
7	109	98.2	109	14	AAR41985	Human T cell react
8	109	98.2	109	20	AAY25593	Felis sp. allergen
9	99	89.2	111	12	AAR12121	TRFP chain 2 - lon
10	92	82.9	92	21	AAB28937	T cell reactive fe
11	92	82.9	92	21	AAY87678	Feline human TRFP
12	92	82.9	92	21	AAY51475	Human TRFP chain 2
13	83	74.8	97	12	AAR12123	TRFP chain 2 - tru
14	83	74.8	109	21	AAB28934	T cell reactive fe
15	83	74.8	109	21	AAY87674	Feline human TRFP
16	83	74.8	109	21	AAY90104	Cat TRFP chain 2 s
17	83	74.8	109	21	AAY51471	Human TRFP chain 2
18	83	74.8	110	13	AAR27370	TRFP Chain #2 with
19	81	73.0	97	21	AAB28935	T cell reactive fe
20	81	73.0	97	21	AAY87675	Feline human TRFP
21	81	73.0	97	21	AAY90105	Cat TRFP chain 2 t
22	81	73.0	97	21	AAY51472	Human TRFP chain 2
23	79	71.2	109	12	AAR12122	TRFP I chain 2 - s
24	66	59.5	92	18	AAW27382	Chain 2 of major c
25	66	59.5	92	20	AAY25532	Human MHC Class II
26	64	57.7	90	21	AAB28938	T cell reactive fe
27	64	57.7	90	21	AAY87679	Feline human TRFP
28	64	57.7	90	21	AAY51476	Human TRFP chain 2
29	51	45.9	101	13	AAR27372	TRFP Chain #2 C2ST
30	37	33.3	51	21	AAB28939	T cell reactive fe
31	37	33.3	51	21	AAY87680	Feline human TRFP
32	37	33.3	51	21	AAY51477	Human TRFP chain 2
33	37	33.3	82	13	AAR27371	TRFP Chain #2 C2ST
34	32	28.8	82	21	AAB28940	T cell reactive fe
35	32	28.8	82	21	AAY87681	Feline human TRFP
36	32	28.8	82	21	AAY90106	Cat TRFP chain 2 g
37	32	28.8	82	21	AAY51478	Human TRFP chain 2
38	26	23.4	26	14	AAR36544	Peptide Z. Felis.
39	26	23.4	26	14	AAR41977	Human T cell react
40	26	23.4	26	21	AAB28943	Peptide Z derived
41	26	23.4	26	21	AAY87684	Feline human TRFP
42	26	23.4	26	21	AAY90109	Cat TRFP derived p
43	26	23.4	26	21	AAY51481	Human TRFP derived
44	26	23.4	96	14	AAR36548	Recombitope YZX.
45	26	23.4	96	21	AAB28979	Peptide YZX. Feli

ALIGNMENTS

RESULT 1
AAR27369
ID AAR27369 standard; protein; 111 AA.
XX
AC AAR27369;
XX
DT 20-MAY-1998 (first entry)
XX
DE TRFP Chain #2 with C2 leader sequence.
XX
KW T cell reactive feline protein; cat allergy; allergic; IgE;
KW desensitizing.
XX
OS Felis domesticus.
XX
FH Key Location/Qualifiers
FT peptide 1..19
FT /label= C2 leader
FT protein 19..111
FT /label= TRFP chain #2
XX
PN WO9215613-A.
XX
PD 17-SEP-1992.
XX
PF 20-FEB-1992; 92WO-US01344.
XX
PR 28-FEB-1991; 91US-0662193.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Bond J, Kuo M;
XX
DR WPI; 1992-331670/40.

```
XX
PT   Modified human T-cell reactive feline protein - stimulates T-cell
PT   in individuals allergic to cats and shows reduced
PT   histamine-releasing properties
XX
PS   Claim 1; Fig 1; 35pp; English.
XX
CC   This sequence represents a modified human T-cell reactive feline
CC   protein which stimulates T-cells from an individual who is allergic
CC   to cats, but which interacts with human IgE to a lesser extent than
CC   does affinity purified TRFP. The protein is modified by treating
CC   with either a mild alkali (pH 12.5-13.5 , KOH, NaOH, LiOH or tertiary
CC   amines) or an enzyme which removes O-linked groups (carbohydrate
CC   moieties). It is useful in desensitising people who are allergic to cats.
XX
SQ   Sequence   111 AA;


  Query Match             100.0%;  Score 111;  DB 13;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-100;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111


RESULT   2
AAB28933
ID   AAB28933 standard; Protein; 111 AA.
XX
AC   AAB28933;
XX
DT   29-JAN-2001  (first entry)
XX
DE   T cell reactive feline protein chain 2 long form.
XX
KW   Cat; allergy; human T cell reactive feline protein; hTRFP;
KW   immunotherapy.
XX
OS   Felis sp.
XX
PN   US6120769-A.
XX
PD   19-SEP-2000.
XX
PF   28-APR-1995;   95US-0431184.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Gefter ML,  Garman RD,  Greenstein JL,  Bond JF;
XX
DR   WPI; 2000-601477/57.
DR   N-PSDB; AAC60103.
XX
PT   Detecting, preventing and treating sensitivity to cat protein allergen
PT   comprises combining a biological sample with a human T cell reactive
PT   feline protein and determining the extent of binding that occurs  -
XX
PS   Claim 1; Figure 3; 106pp; English.
XX
CC   The present invention relates to the detection of sensitivity to a cat
CC   protein allergen by combining a blood sample from a subject with a
CC   peptide of human T cell reactive feline protein (hTRFP). This method
CC   and the hTRFP peptides are useful for diagnosing, preventing and
CC   treating cat allergies by reducing or abolishing an individual's
CC   allergic response to a cat allergen. DNA encoding the TRFP may be
CC   used as probes to locate equivalent sequences present in other species.
CC   These may further be used to study the mechanism of immunotherapy of
CC   cat allergy, and to design modified derivatives, analogues or
CC   functional equivalents useful in immunotherapy. The present
CC   sequence was used in the invention.
XX
SQ   Sequence   111 AA;


  Query Match             100.0%;  Score 111;  DB 21;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-100;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111


RESULT   3
AAY87673
ID   AAY87673 standard; Protein; 111 AA.
XX
AC   AAY87673;
XX
DT   22-AUG-2000  (first entry)
XX
DE   Feline human TRFP chain 2 long form protein.
XX
KW   T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
KW   antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX
OS   Felis sp.
XX
PN   US6048962-A.
XX
PD   11-APR-2000.
XX
PF   27-APR-1995;   95US-0430014.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Kuo M,  Rogers BL,  Gefter ML,  Morgenstern JP,  Brauer AW;
PI   Greenstein JL,  Griffith IJ,  Garman RD;
XX
DR   WPI; 2000-316905/27.
DR   N-PSDB; AAA12244.
XX
PT   New human T cell reactive feline protein useful for reducing or
PT   abolishing individual's allergic response to cat allergen comprising
PT   two different covalently linked peptide chains  -
XX
PS   Claim 2; Column 75-76; 106pp; English.
XX
CC   This invention describes a novel naturally occurring cat protein allergen
CC   (I), human T cell reactive feline protein (TRFP), comprising two
```

```
CC   different covalently linked peptide chains with a molecular weight of 20
CC   kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
CC   under reducing conditions. The products of the invention have
CC   antiallergic activity and act as human T cell stimulators. TRFP is useful
CC   for reducing or preventing the adverse effects of cat allergens on cat
CC   allergic individuals and in ex vivo diagnostic tests to determine which
CC   peptides cause sensitivity so as to selectively use them to desensitize
CC   a cat sensitive individual. Purified TRFP is also useful for studying
CC   the mechanism of immunotherapy of cat allergy and to design modified
CC   derivatives, analogs or functional equivalents that are more useful in
CC   immunotherapy against cat allergy. DNA sequences encoding TRFP are
CC   useful as probes to locate equivalent sequences present in other species
CC   (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC   and/or therapeutics. Fully defined and characterized TRFP provides
CC   complete and a very simple desensitization therapy. This sequence
CC   represents a human T cell reactive feline protein (also known as Fel d I)
CC   chain 2, long form which is described in the method of the invention.
XX
SQ   Sequence   111 AA;


  Query Match             100.0%;  Score 111;  DB 21;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-100;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111


RESULT   4
AAY90103
ID   AAY90103 standard; Protein; 111 AA.
XX
AC   AAY90103;
XX
DT   13-JUL-2000  (first entry)
XX
DE   Cat TRFP chain 2 long form protein sequence.
XX
KW   Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW   house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW   diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS   Felis domesticus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..111
FT                   /note= "mature TRFP chain 2 long form"
XX
PN   US6025162-A.
XX
PD   15-FEB-2000.
XX
PF   28-APR-1995;   95US-0430944.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Morgenstern JP,  Griffith IJ,  Rogers BL;
XX
DR   WPI; 2000-181812/16.
DR   N-PSDB; AAA07437.
XX
PT   New human T cell reactive feline protein, useful for desensitizing cat
PT   allergic individuals to cat allergens -
XX
PS   Claim 1; Fig 3; 108pp; English.
XX
CC   This sequence is a peptide chain of the human T cell reactive feline
CC   protein (TRFP) of the invention. The protein is a cat protein allergen,
CC   and was isolated from a vacuum bag extract obtained by affinity
CC   purification of house dust collected from several homes with cats. TRFP
CC   is composed of two covalently linked peptide chains, and is also referred
CC   to as Fel d I. TRFP and its peptides are useful for reducing or
CC   preventing the adverse effects that exposure to cat allergens normally
CC   has on cat allergic individuals (i.e. to desensitise individuals to cat
CC   allergens or block the effect of the allergens). TRFP is also used in
CC   methods of diagnosing sensitivity to Felis domesticus in an individual.
CC   DNA sequences encoding TRFP can be used as probes to locate equivalent
CC   sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC   dog, that may be useful in a diagnostic and/or therapeutic applications.
XX
SQ   Sequence   111 AA;


  Query Match             100.0%;  Score 111;  DB 21;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-100;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111


RESULT   5
AAY51470
ID   AAY51470 standard; Protein; 111 AA.
XX
AC   AAY51470;
XX
DT   22-MAY-2000  (first entry)
XX
DE   Human TRFP chain 2 (long form) protein fragment.
XX
KW   T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;
KW   down regulation; immune response; allergen; immunoglobulin E;
KW   sensitivity; cat protein allergen; human; chain 2.
XX
OS   Homo sapiens.
XX
PN   US6019972-A.
XX
PD   01-FEB-2000.
XX
PF   02-SEP-1994;   94US-0300928.
XX
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Garman RD,  Greenstein JL,  Kuo M,  Briner TJ,  Morville M,  Gefter ML;
XX
DR   WPI; 2000-146862/13.
```

```
DR   N-PSDB; AAZ88617.
XX
PT   Peptides of human T cell reactive feline protein for treating
PT   sensitivity to cat protein allergens comprise at least one T cell
PT   epitope recognized by a T cell receptor specific for the human T cell
PT   reactive feline protein -
XX
PS   Claim 1; Column 75-78; 105pp; English.
XX
CC   This invention describes a novel peptide (I) of human T cell reactive
CC   feline protein (hTRFP) having at least one T cell epitope recognized
CC   by a T cell receptor specific for the human T cell reactive feline
CC   protein, the peptide consisting of at least 7-30 amino acids, and having
CC   an amino acid sequence derived from an amino acid sequence comprising 94,
CC   96, 97, 109, or 111 residues, given in the specification. The peptides
CC   down regulate the immune response to the allergen. The peptides have
CC   reduced immunoglobulin E binding and reduce T cell responsiveness. The
CC   peptide (I) is useful in compositions for treating sensitivity to a cat
CC   protein allergen in a subject. This sequence represents the human TRFP
CC   chain 2 (long form).
XX
SQ   Sequence   111 AA;


  Query Match             100.0%;  Score 111;  DB 21;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-100;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertа 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111


RESULT   6
AAR36541
ID   AAR36541 standard; Protein; 109 AA.
XX
AC   AAR36541;
XX
DT   12-AUG-1993  (first entry)
XX
DE   TRFP chain 2 (with Leader).
XX
KW   Human T cell reactive feline protein; TRFP; leader A; leader B;
KW   epitope.
XX
OS   Felis.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..17
FT                   /label= leader_peptide
XX
PN   WO9308280-A.
XX
PD   29-APR-1993.
XX
PF   16-OCT-1992;   92WO-US08694.
XX
PR   16-OCT-1991;   91US-0777859.
PR   13-DEC-1991;   91US-0807529.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Bond JF,  Garman RD,  Kuo M,  Morgenstern JP,  Morville M;
PI   Rogers BL;
XX
DR   WPI; 1993-152473/18.
DR   N-PSDB; AAQ41558.
XX
PT   Recombitope peptide having T-cell stimulating activity - for the
PT   diagnosis and treatment of sensitivity to protein allergens;
PT   auto:antigens and protein antigens
XX
PS   Disclosure; Fig 2; 73pp; English.
XX
CC   Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
CC   coli and purified.  T cell epitope studies using overlapping peptide
CC   regions derived from the TRFP amino acids sequence were used to
CC   identify multiple T cell epitopes in each chain of TRFP.
XX
SQ   Sequence   109 AA;


  Query Match             98.2%;  Score 109;  DB 14;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-98;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertamk 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 kiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 109


RESULT   7
AAR41985
ID   AAR41985 standard; Protein; 109 AA.
XX
AC   AAR41985;
XX
DT   21-APR-1994  (first entry)
XX
DE   Human T cell reactive feline protein chain 2.
XX
KW   Human; T cell; reactive; feline; protein; immune response; antigen;
KW   tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW   Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW   Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..17
FT                   /note= "Signal peptide"
FT   Protein         18..109
FT                   /note= "Mature protein"
XX
PN   WO9319178-A.
XX
PD   30-SEP-1993.
XX
PF   25-MAR-1993;   93WO-US02462.
XX
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI   Briner TJ,  Garman RD,  Gefter ML,  Greenstein JL,  Kuo M;
PI   Morville M;
XX
DR   WPI; 1993-320744/40.
DR   N-PSDB; AAQ49535.
XX
PT   New peptide(s) for inducing tolerance - comprise one or more
PT   epitope(s) of an allergen administered subcutaneously, for
PT   treating sensitivity to cats, bees, etc.
XX
PS   Disclosure; Fig 2; 107pp; English.
```

```
XX
CC   This sequence represents chain 2 of human T cell reactive feline
CC   protein (TRFP).  Peptides derived from TRFP may be used in a
CC   therapeutic composition which is useful in treating diseases which
CC   involve an immune response to a protein antigen.  This composition
CC   may be used to induce tolerance in a mammal to Dermatophagoides,
CC   Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
CC   Quercus, Olea, Artemesia, Plantago, Parietaria, Canis, Blattella,
CC   Apis, Periplaneta and to autoantigens in humans.
XX
SQ   Sequence   109 AA;


  Query Match             98.2%;  Score 109;  DB 14;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-98;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertamk 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 kiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 109


RESULT   8
AAY25593
ID   AAY25593 standard; protein; 109 AA.
XX
AC   AAY25593;
XX
DT   30-SEP-1999  (first entry)
XX
DE   Felis sp. allergen 1082946 Fel dI chain 2 precursor protein fragment.
XX
KW   Major histocompatibility complex; class II; desensitising; human;
KW   allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;
KW   chiromidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW   screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW   cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW   mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS   Felis sp.
XX
PN   WO9934826-A1.
XX
PD   15-JUL-1999.
XX
PF   11-JAN-1999;   99WO-GB00080.
XX
PR   21-SEP-1998;   98GB-0020474.
PR   09-JAN-1998;   98GB-0000445.
XX
PA   (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI   Kay AB,  Larche M;
XX
DR   WPI; 1999-458255/38.
XX
PT   Desensitizing patients to polypeptide allergens
XX
PS   Example 6; Page 53; 117pp; English.
XX
CC   This invention describes a novel method of desensitizing a patient to a
CC   polypeptide allergen and comprises administering to the patient a peptide
CC   derived from the allergen where restriction to a MHC Class II molecule
CC   possessed by the patient can be demonstrated for the peptide and the
CC   peptide is able to induce a late phase response in an individual who
CC   possesses the MHC Class II molecule. The methods can be used for
CC   desensitising patients to allergens present in e.g. grass, tree and weed
CC   (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC   the chiromidae (non-biting midges), spiders and mites, housefly, fruit
CC   fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC   non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC   Tenibrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC   sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC   produce immunological vaccines which may be used to prevent and/or treat
CC   conditions involving hypersensitivity to allergens. This sequence
CC   represents the cat allergen 1082946 Fel dI chain 2 precursor.
XX
SQ   Sequence   109 AA;


  Query Match             98.2%;  Score 109;  DB 20;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-98;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertamk 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 kiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 109


RESULT   9
AAR12121
ID   AAR12121 standard; Protein; 111 AA.
XX
AC   AAR12121;
XX
DT   26-JUL-1991  (first entry)
XX
DE   TRFP chain 2 - long form.
XX
KW   Human T cell reactive feline protein; cat allergens.
XX
OS   Felis catus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         3..19
FT                   /label= Leader sequence
FT   Protein         20..111
FT                   /label= TRFP I long form
XX
PN   WO9106571-A.
XX
PD   16-MAY-1991.
XX
PF   02-NOV-1990;   90WO-US06548.
XX
PR   03-NOV-1989;   89US-0431565.
XX
PA   (IMMU-) IMMULOGIC PHARM COR.
XX
PI   Gefter ML,  Garman RD,  Greenstein JL,  Juo M,  Rogers BL;
PI   Brauer AW;
XX
DR   WPI; 1991-164136/22.
DR   N-PSDB; AAQ11838.
XX
PT   New pure covalently linked human T cell reactive feline protein -
PT   and modified peptide(s), used to reduce effects of cat allergens
PT   and to diagnose sensitivity to allergens.
XX
PS   Claim 2; Fig 3; 70pp; English.
XX
CC   Poly-A mRNA from cat parotid and mandibular glands was used to
CC   produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC   clones were then used to screen a cat genomic library. Chain 1
CC   exists in two forms having different leader sequences (A and B).
CC   The sequence can be used to express the protein and peptide derivs.
CC   which stimulate T-cells in persons allergic to cats.  The peptides
CC   can be used to reduce/eliminate the allergic response partic. by
```

CC modificn. of lynphokine prodn. by the T-cells. They can also be
CC used to identify epitopes responsible for sensitivity. The DNA can
CC be used to detect comparable sequence in other species, and also
CC for prodn. of modified forms of TRFP esp. showing reduced binding
CC to IgE and thus reduced tendency to cause adverse reactions.
CC See also AAR12119-R12123.
XX
SQ Sequence 111 AA;

Query Match 89.2%; Score 99; DB 12; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.2e-88;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQN 99
|||||||||||||||||||||||||||||||||||||||
Db 61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqn 99

RESULT 10
AAB28937
ID AAB28937 standard; Protein; 92 AA.
XX
AC AAB28937;
XX
DT 29-JAN-2001 (first entry)
XX
DE T cell reactive feline protein chain 2 PRO long.
XX
KW Cat; allergy; human T cell reactive feline protein; hTRFP;
KW immunotherapy.
XX
OS Felis sp.
XX
PN US6120769-A.
XX
PD 19-SEP-2000.
XX
PF 28-APR-1995; 95US-0431184.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Gefter ML, Garman RD, Greenstein JL, Bond JF;
XX
DR WPI; 2000-601477/57.
XX
PT Detecting, preventing and treating sensitivity to cat protein allergen
PT comprises combining a biological sample with a human T cell reactive
PT feline protein and determining the extent of binding that occurs -
XX
PS Disclosure; Figure 7; 106pp; English.
XX
CC The present invention relates to the detection of sensitivity to a cat
CC protein allergen by combining a blood sample from a subject with a
CC peptide of human T cell reactive feline protein (hTRFP). This method
CC and the hTRFP peptides are useful for diagnosing, preventing and
CC treating cat allergies by reducing or abolishing an individual's
CC allergic response to a cat allergen. DNA encoding the TRFP may be
CC used as probes to locate equivalent sequences present in other species.
CC These may further be used to study the mechanism of immunotherapy of
CC cat allergy, and to design modified derivatives, analogues or
CC functional equivalents useful in immunotherapy. The present
CC sequence was used in the invention.
XX
SQ Sequence 92 AA;

Query Match 82.9%; Score 92; DB 21; Length 92;
Best Local Similarity 100.0%; Pred. No. 6.1e-82;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 vkmaetcpifydvffavangnellldlsltkvnatepertamkkiqdcyvenglisrvld 60

Qy 80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
||||||||||||||||||||||||||||||||
Db 61 glvmttissskdcmgeavqntvedlklntlgr 92

RESULT 11
AAY87678
ID AAY87678 standard; Protein; 92 AA.
XX
AC AAY87678;
XX
DT 22-AUG-2000 (first entry)
XX
DE Feline human TRFP chain 2 long form protein #2.
XX
KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
KW antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX
OS Felis sp.
XX
PN US6048962-A.
XX
PD 11-APR-2000.
XX
PF 27-APR-1995; 95US-0430014.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Kuo M, Rogers BL, Gefter ML, Morgenstern JP, Brauer AW;
PI Greenstein JL, Griffith IJ, Garman RD;
XX
DR WPI; 2000-316905/27.
XX
PT New human T cell reactive feline protein useful for reducing or
PT abolishing individual's allergic response to cat allergen comprising
PT two different covalently linked peptide chains -
XX
PS Example 2; Column 81-84; 106pp; English.
XX
CC This invention describes a novel naturally occurring cat protein allergen
CC (I), human T cell reactive feline protein (TRFP), comprising two
CC different covalently linked peptide chains with a molecular weight of 20
CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
CC under reducing conditions. The products of the invention have
CC antiallergic activity and act as human T cell stimulators. TRFP is useful
CC for reducing or preventing the adverse effects of cat allergens on cat
CC allergic individuals and in ex vivo diagnostic tests to determine which
CC peptides cause sensitivity so as to selectively use them to desensitize
CC a cat sensitive individual. Purified TRFP is also useful for studying
CC the mechanism of immunotherapy of cat allergy and to design modified
CC derivatives, analogs or functional equivalents that are more useful in

```
CC   immunotherapy against cat allergy. DNA sequences encoding TRFP are
CC   useful as probes to locate equivalent sequences present in other species
CC   (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC   and/or therapeutics. Fully defined and characterized TRFP provides
CC   complete and a very simple desensitization therapy. This sequence
CC   represents a human T cell reactive feline protein (also known as Fel d I)
CC   chain 2, long form which is described in the method of the invention.
XX
SQ   Sequence   92 AA;


  Query Match             82.9%;  Score 92;  DB 21;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-82;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vkmaetcpifydvffavangnellldlsltkvnatepertamkkiqdcyvenglisrvld 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 glvmttissskdcmgeavqntvedlklntlgr 92


RESULT  12
AAY51475
ID   AAY51475 standard; Protein; 92 AA.
XX
AC   AAY51475;
XX
DT   22-MAY-2000  (first entry)
XX
DE   Human TRFP chain 2 (long form) protein fragment #2.
XX
KW   T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;
KW   down regulation; immune response; allergen; immunoglobulin E;
KW   sensitivity; cat protein allergen; human; chain 2.
XX
OS   Homo sapiens.
XX
PN   US6019972-A.
XX
PD   01-FEB-2000.
XX
PF   02-SEP-1994;   94US-0300928.
XX
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Garman RD,  Greenstein JL,  Kuo M,  Briner TJ,  Morville M,  Gefter ML;
XX
DR   WPI; 2000-146862/13.
XX
PT   Peptides of human T cell reactive feline protein for treating
PT   sensitivity to cat protein allergens comprise at least one T cell
PT   epitope recognized by a T cell receptor specific for the human T cell
PT   reactive feline protein -
XX
PS   Example 1; Column 83-84; 105pp; English.
XX
CC   This invention describes a novel peptide (I) of human T cell reactive
CC   feline protein (hTRFP) having at least one T cell epitope recognized
CC   by a T cell receptor specific for the human T cell reactive feline
CC   protein, the peptide consisting of at least 7-30 amino acids, and having
CC   an amino acid sequence derived from an amino acid sequence comprising 94,
CC   96. 97, 109, or 111 residues, given in the specification. The peptides
CC   down regulate the immune response to the allergen. The peptides have
CC   reduced immunoglobulin E binding and reduce T cell responsiveness. The
CC   peptide (I) is useful in compositions for treating sensitivity to a cat
CC   protein allergen in a subject. This sequence represents the human TRFP
CC   chain 2 (long form).
XX
SQ   Sequence   92 AA;


  Query Match             82.9%;  Score 92;  DB 21;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-82;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vkmaetcpifydvffavangnellldlsltkvnatepertamkkiqdcyvenglisrvld 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 glvmttissskdcmgeavqntvedlklntlgr 92


RESULT  13
AAR12123
ID   AAR12123 standard; Protein; 97 AA.
XX
AC   AAR12123;
XX
DT   26-JUL-1991  (first entry)
XX
DE   TRFP chain 2 - truncated short form.
XX
KW   Human T cell reactive feline protein; cat allergens.
XX
OS   Felis catus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         3..19
FT                   /label= Leader sequence
FT   Protein         20..97
FT                   /label= TRFP I truncated short form
XX
PN   WO9106571-A.
XX
PD   16-MAY-1991.
XX
PF   02-NOV-1990;   90WO-US06548.
XX
PR   03-NOV-1989;   89US-0431565.
XX
PA   (IMMU-) IMMULOGIC PHARM COR.
XX
PI   Gefter ML,  Garman RD,  Greenstein JL,  Juo M,  Rogers BL;
PI   Brauer AW;
XX
DR   WPI; 1991-164136/22.
DR   N-PSDB; AAQ11840.
XX
PT   New pure covalently linked human T cell reactive feline protein -
PT   and modified peptide(s), used to reduce effects of cat allergens
PT   and to diagnose sensitivity to allergens.
XX
PS   Claim 2; Fig 5; 70pp; English.
XX
CC   Poly-A mRNA from cat parotid and mandibular glands was used to
CC   produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC   clones were then used to screen a cat genomic library. Chain 1
CC   exists in two forms having different leader sequences (A and B).
CC   The sequence can be used to express the protein and peptide derivs.
CC   which stimulate T-cells in persons allergic to cats.  The peptides
CC   can be used to reduce/eliminate the allergic response partic. by
CC   modificn. of lynphokine prodn. by the T-cells.  They can also be
CC   used to identify epitopes responsible for sensitivity.  The DNA can
CC   be used to detect comparable sequence in other species, and also
```

```
CC   for prodn. of modified forms of TRFP esp. showing reduced binding
CC   to IgE and thus reduced tendency to cause adverse reactions.
CC   See also AAR12119-R12122.
XX
SQ   Sequence   97 AA;


  Query Match             74.8%;  Score 83;  DB 12;  Length 97;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-73;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvm 83


RESULT  14
AAB28934
ID   AAB28934 standard; Protein; 109 AA.
XX
AC   AAB28934;
XX
DT   29-JAN-2001  (first entry)
XX
DE   T cell reactive feline protein chain 2 short form.
XX
KW   Cat; allergy; human T cell reactive feline protein; hTRFP;
KW   immunotherapy.
XX
OS   Felis sp.
XX
PN   US6120769-A.
XX
PD   19-SEP-2000.
XX
PF   28-APR-1995;   95US-0431184.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Gefter ML,  Garman RD,  Greenstein JL,  Bond JF;
XX
DR   WPI; 2000-601477/57.
DR   N-PSDB; AAC60104.
XX
PT   Detecting, preventing and treating sensitivity to cat protein allergen
PT   comprises combining a biological sample with a human T cell reactive
PT   feline protein and determining the extent of binding that occurs  -
XX
PS   Claim ; Figure 4; 106pp; English.
XX
CC   The present invention relates to the detection of sensitivity to a cat
CC   protein allergen by combining a blood sample from a subject with a
CC   peptide of human T cell reactive feline protein (hTRFP). This method
CC   and the hTRFP peptides are useful for diagnosing, preventing and
CC   treating cat allergies by reducing or abolishing an individual's
CC   allergic response to a cat allergen. DNA encoding the TRFP may be
CC   used as probes to locate equivalent sequences present in other species.
CC   These may further be used to study the mechanism of immunotherapy of
CC   cat allergy, and to design modified derivatives, analogues or
CC   functional equivalents useful in immunotherapy. The present
CC   sequence was used in the invention.
XX
SQ   Sequence   109 AA;


  Query Match             74.8%;  Score 83;  DB 21;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-73;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvm 83


RESULT  15
AAY87674
ID   AAY87674 standard; Protein; 109 AA.
XX
AC   AAY87674;
XX
DT   22-AUG-2000  (first entry)
XX
DE   Feline human TRFP chain 2 short form protein.
XX
KW   T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
KW   antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX
OS   Felis sp.
XX
PN   US6048962-A.
XX
PD   11-APR-2000.
XX
PF   27-APR-1995;   95US-0430014.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Kuo M,  Rogers BL,  Gefter ML,  Morgenstern JP,  Brauer AW;
PI   Greenstein JL,  Griffith IJ,  Garman RD;
XX
DR   WPI; 2000-316905/27.
DR   N-PSDB; AAA12245.
XX
PT   New human T cell reactive feline protein useful for reducing or
PT   abolishing individual's allergic response to cat allergen comprising
PT   two different covalently linked peptide chains  -
XX
PS   Claim 2; Column 77-78; 106pp; English.
XX
CC   This invention describes a novel naturally occurring cat protein allergen
CC   (I), human T cell reactive feline protein (TRFP), comprising two
CC   different covalently linked peptide chains with a molecular weight of 20
CC   kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
CC   under reducing conditions. The products of the invention have
CC   antiallergic activity and act as human T cell stimulators. TRFP is useful
CC   for reducing or preventing the adverse effects of cat allergens on cat
CC   allergic individuals and in ex vivo diagnostic tests to determine which
CC   peptides cause sensitivity so as to selectively use them to desensitize
CC   a cat sensitive individual. Purified TRFP is also useful for studying
CC   the mechanism of immunotherapy of cat allergy and to design modified
CC   derivatives, analogs or functional equivalents that are more useful in
CC   immunotherapy against cat allergy. DNA sequences encoding TRFP are
```

CC useful as probes to locate equivalent sequences present in other species
CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC and/or therapeutics. Fully defined and characterized TRFP provides
CC complete and a very simple desensitization therapy. This sequence
CC represents a human T cell reactive feline protein (also known as Fel d I)
CC chain 2, short form which is described in the method of the invention.
XX
SQ Sequence 109 AA;

Query Match 74.8%; Score 83; DB 21; Length 109;
Best Local Similarity 100.0%; Pred. No. 4.1e-73;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvm 83
```

Search completed: May 7, 2002, 12:17:34
Job time: 201 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:10:08 ; Search time 23.73 Seconds
(without alignments)
346.487 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 552
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	552	100.0	111	13	AAR27369	TRFP Chain #2 with
2	552	100.0	111	21	AAB28933	T cell reactive fe
3	552	100.0	111	21	AAY87673	Feline human TRFP
4	552	100.0	111	21	AAY90103	Cat TRFP chain 2 l
5	552	100.0	111	21	AAY51470	Human TRFP chain 2
6	547	99.1	111	12	AAR12121	TRFP chain 2 - lon
7	541	98.0	109	14	AAR36541	TRFP chain 2 (with
8	541	98.0	109	14	AAR41985	Human T cell react
9	541	98.0	109	20	AAY25593	Felis sp. allergen
10	512.5	92.8	110	13	AAR27370	TRFP Chain #2 with
11	506	91.7	109	21	AAB28934	T cell reactive fe
12	506	91.7	109	21	AAY87674	Feline human TRFP
13	506	91.7	109	21	AAY90104	Cat TRFP chain 2 s
14	506	91.7	109	21	AAY51471	Human TRFP chain 2
15	498	90.2	109	12	AAR12122	TRFP I chain 2 - s
16	465	84.2	92	21	AAB28937	T cell reactive fe
17	465	84.2	92	21	AAY87678	Feline human TRFP
18	465	84.2	92	21	AAY51475	Human TRFP chain 2
19	458	83.0	92	18	AAW27382	Chain 2 of major c
20	458	83.0	92	20	AAY25532	Human MHC Class II
21	457.5	82.9	101	13	AAR27372	TRFP Chain #2 C2ST
22	419	75.9	90	21	AAB28938	T cell reactive fe
23	419	75.9	90	21	AAY87679	Feline human TRFP
24	419	75.9	90	21	AAY51476	Human TRFP chain 2
25	412	74.6	97	12	AAR12123	TRFP chain 2 - tru
26	408	73.9	97	21	AAB28935	T cell reactive fe
27	408	73.9	97	21	AAY87675	Feline human TRFP
28	408	73.9	97	21	AAY90105	Cat TRFP chain 2 t
29	408	73.9	97	21	AAY51472	Human TRFP chain 2
30	370.5	67.1	82	21	AAB28940	T cell reactive fe
31	370.5	67.1	82	21	AAY87681	Feline human TRFP
32	370.5	67.1	82	21	AAY90106	Cat TRFP chain 2 g
33	370.5	67.1	82	21	AAY51478	Human TRFP chain 2
34	288.5	52.3	82	13	AAR27371	TRFP Chain #2 C2ST
35	185	33.5	51	21	AAB28939	T cell reactive fe
36	185	33.5	51	21	AAY87680	Feline human TRFP
37	185	33.5	51	21	AAY51477	Human TRFP chain 2
38	128.5	23.3	96	14	AAR36548	Recombitope YZX.
39	128.5	23.3	96	21	AAB28979	Peptide YZX. Feli
40	128.5	23.3	96	21	AAY90145	TRFP fusion protei
41	128	23.2	26	14	AAR36544	Peptide Z. Felis.
42	128	23.2	26	14	AAR41977	Human T cell react
43	128	23.2	26	21	AAB28943	Peptide Z derived
44	128	23.2	26	21	AAY87684	Feline human TRFP
45	128	23.2	26	21	AAY90109	Cat TRFP derived p

ALIGNMENTS

RESULT 1
AAR27369
ID AAR27369 standard; protein; 111 AA.
XX
AC AAR27369;
XX
DT 20-MAY-1998 (first entry)
XX
DE TRFP Chain #2 with C2 leader sequence.
XX
KW T cell reactive feline protein; cat allergy; allergic; IgE;
KW desensitizing.
XX
OS Felis domesticus.
XX
FH Key Location/Qualifiers
FT peptide 1..19
FT /label= C2 leader
FT protein 19..111
FT /label= TRFP chain #2
XX
PN WO9215613-A.
XX
PD 17-SEP-1992.
XX
PF 20-FEB-1992; 92WO-US01344.
XX
PR 28-FEB-1991; 91US-0662193.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Bond J, Kuo M;
XX
DR WPI; 1992-331670/40.

XX
PT Modified human T-cell reactive feline protein - stimulates T-cell
PT in individuals allergic to cats and shows reduced
PT histamine-releasing properties
XX
PS Claim 1; Fig 1; 35pp; English.
XX
CC This sequence represents a modified human T-cell reactive feline
CC protein which stimulates T-cells from an individual who is allergic
CC to cats, but which interacts with human IgE to a lesser extent than
CC does affinity purified TRFP. The protein is modified by treating
CC with either a mild alkali (pH 12.5-13.5 , KOH, NaOH, LiOH or tertiary
CC amines) or an enzyme which removes O-linked groups (carbohydrate
CC moieties). It is useful in desensitising people who are allergic to cats.
XX
SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 13; Length 111;
Best Local Similarity 100.0%; Pred. No. 2e-57;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111

RESULT 2
AAB28933
ID AAB28933 standard; Protein; 111 AA.
XX
AC AAB28933;
XX
DT 29-JAN-2001 (first entry)
XX
DE T cell reactive feline protein chain 2 long form.
XX
KW Cat; allergy; human T cell reactive feline protein; hTRFP;
KW immunotherapy.
XX
OS Felis sp.
XX
PN US6120769-A.
XX
PD 19-SEP-2000.
XX
PF 28-APR-1995; 95US-0431184.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Gefter ML, Garman RD, Greenstein JL, Bond JF;
XX
DR WPI; 2000-601477/57.
DR N-PSDB; AAC60103.
XX
PT Detecting, preventing and treating sensitivity to cat protein allergen
PT comprises combining a biological sample with a human T cell reactive
PT feline protein and determining the extent of binding that occurs -
XX
PS Claim 1; Figure 3; 106pp; English.
XX
CC The present invention relates to the detection of sensitivity to a cat
CC protein allergen by combining a blood sample from a subject with a
CC peptide of human T cell reactive feline protein (hTRFP). This method
CC and the hTRFP peptides are useful for diagnosing, preventing and
CC treating cat allergies by reducing or abolishing an individual's
CC allergic response to a cat allergen. DNA encoding the TRFP may be
CC used as probes to locate equivalent sequences present in other species.
CC These may further be used to study the mechanism of immunotherapy of
CC cat allergy, and to design modified derivatives, analogues or
CC functional equivalents useful in immunotherapy. The present
CC sequence was used in the invention.
XX
SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 21; Length 111;
Best Local Similarity 100.0%; Pred. No. 2e-57;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111

RESULT 3
AAY87673
ID AAY87673 standard; Protein; 111 AA.
XX
AC AAY87673;
XX
DT 22-AUG-2000 (first entry)
XX
DE Feline human TRFP chain 2 long form protein.
XX
KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
KW antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX
OS Felis sp.
XX
PN US6048962-A.
XX
PD 11-APR-2000.
XX
PF 27-APR-1995; 95US-0430014.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Kuo M, Rogers BL, Gefter ML, Morgenstern JP, Brauer AW;
PI Greenstein JL, Griffith IJ, Garman RD;
XX
DR WPI; 2000-316905/27.
DR N-PSDB; AAA12244.
XX
PT New human T cell reactive feline protein useful for reducing or
PT abolishing individual's allergic response to cat allergen comprising
PT two different covalently linked peptide chains -
XX
PS Claim 2; Column 75-76; 106pp; English.
XX
CC This invention describes a novel naturally occurring cat protein allergen
CC (I), human T cell reactive feline protein (TRFP), comprising two

CC different covalently linked peptide chains with a molecular weight of 20
CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
CC under reducing conditions. The products of the invention have
CC antiallergic activity and act as human T cell stimulators. TRFP is useful
CC for reducing or preventing the adverse effects of cat allergens on cat
CC allergic individuals and in ex vivo diagnostic tests to determine which
CC peptides cause sensitivity so as to selectively use them to desensitize
CC a cat sensitive individual. Purified TRFP is also useful for studying
CC the mechanism of immunotherapy of cat allergy and to design modified
CC derivatives, analogs or functional equivalents that are more useful in
CC immunotherapy against cat allergy. DNA sequences encoding TRFP are
CC useful as probes to locate equivalent sequences present in other species
CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC and/or therapeutics. Fully defined and characterized TRFP provides
CC complete and a very simple desensitization therapy. This sequence
CC represents a human T cell reactive feline protein (also known as Fel d I)
CC chain 2, long form which is described in the method of the invention.
XX
SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 21; Length 111;
Best Local Similarity 100.0%; Pred. No. 2e-57;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
Db 61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111

RESULT 4
AAY90103
ID AAY90103 standard; Protein; 111 AA.
XX
AC AAY90103;
XX
DT 13-JUL-2000 (first entry)
XX
DE Cat TRFP chain 2 long form protein sequence.
XX
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS Felis domesticus.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /note= "signal peptide"
FT Protein 20..111
FT /note= "mature TRFP chain 2 long form"
XX
PN US6025162-A.
XX
PD 15-FEB-2000.
XX
PF 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Morgenstern JP, Griffith IJ, Rogers BL;
XX
DR WPI; 2000-181812/16.
DR N-PSDB; AAA07437.
XX
PT New human T cell reactive feline protein, useful for desensitizing cat
PT allergic individuals to cat allergens -
XX
PS Claim 1; Fig 3; 108pp; English.
XX
CC This sequence is a peptide chain of the human T cell reactive feline
CC protein (TRFP) of the invention. The protein is a cat protein allergen,
CC and was isolated from a vacuum bag extract obtained by affinity
CC purification of house dust collected from several homes with cats. TRFP
CC is composed of two covalently linked peptide chains, and is also referred
CC to as Fel d I. TRFP and its peptides are useful for reducing or
CC preventing the adverse effects that exposure to cat allergens normally
CC has on cat allergic individuals (i.e. to desensitise individuals to cat
CC allergens or block the effect of the allergens). TRFP is also used in
CC methods of diagnosing sensitivity to Felis domesticus in an individual.
CC DNA sequences encoding TRFP can be used as probes to locate equivalent
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC dog, that may be useful in a diagnostic and/or therapeutic applications.
XX
SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 21; Length 111;
Best Local Similarity 100.0%; Pred. No. 2e-57;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
Db 61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111

RESULT 5
AAY51470
ID AAY51470 standard; Protein; 111 AA.
XX
AC AAY51470;
XX
DT 22-MAY-2000 (first entry)
XX
DE Human TRFP chain 2 (long form) protein fragment.
XX
KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;
KW down regulation; immune response; allergen; immunoglobulin E;
KW sensitivity; cat protein allergen; human; chain 2.
XX
OS Homo sapiens.
XX
PN US6019972-A.
XX
PD 01-FEB-2000.
XX
PF 02-SEP-1994; 94US-0300928.
XX
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gefter ML;
XX
DR WPI; 2000-146862/13.

```
DR   N-PSDB; AAZ88617.
XX
PT   Peptides of human T cell reactive feline protein for treating
PT   sensitivity to cat protein allergens comprise at least one T cell
PT   epitope recognized by a T cell receptor specific for the human T cell
PT   reactive feline protein -
XX
PS   Claim 1; Column 75-78; 105pp; English.
XX
CC   This invention describes a novel peptide (I) of human T cell reactive
CC   feline protein (hTRFP) having at least one T cell epitope recognized
CC   by a T cell receptor specific for the human T cell reactive feline
CC   protein, the peptide consisting of at least 7-30 amino acids, and having
CC   an amino acid sequence derived from an amino acid sequence comprising 94,
CC   96, 97, 109, or 111 residues, given in the specification. The peptides
CC   down regulate the immune response to the allergen. The peptides have
CC   reduced immunoglobulin E binding and reduce T cell responsiveness. The
CC   peptide (I) is useful in compositions for treating sensitivity to a cat
CC   protein allergen in a subject. This sequence represents the human TRFP
CC   chain 2 (long form).
XX
SQ   Sequence   111 AA;


  Query Match             100.0%;  Score 552;  DB 21;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2e-57;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 111


RESULT   6
AAR12121
ID   AAR12121 standard; Protein; 111 AA.
XX
AC   AAR12121;
XX
DT   26-JUL-1991  (first entry)
XX
DE   TRFP chain 2 - long form.
XX
KW   Human T cell reactive feline protein; cat allergens.
XX
OS   Felis catus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         3..19
FT                   /label= Leader sequence
FT   Protein         20..111
FT                   /label= TRFP I long form
XX
PN   WO9106571-A.
XX
PD   16-MAY-1991.
XX
PF   02-NOV-1990;   90WO-US06548.
XX
PR   03-NOV-1989;   89US-0431565.
XX
PA   (IMMU-) IMMULOGIC PHARM COR.
XX
PI   Gefter ML,  Garman RD,  Greenstein JL,  Juo M,  Rogers BL;
PI   Brauer AW;
XX
DR   WPI; 1991-164136/22.
DR   N-PSDB; AAQ11838.
XX
PT   New pure covalently linked human T cell reactive feline protein -
PT   and modified peptide(s), used to reduce effects of cat allergens
PT   and to diagnose sensitivity to allergens.
XX
PS   Claim 2; Fig 3; 70pp; English.
XX
CC   Poly-A mRNA from cat parotid and mandibular glands was used to
CC   produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC   clones were then used to screen a cat genomic library. Chain 1
CC   exists in two forms having different leader sequences (A and B).
CC   The sequence can be used to express the protein and peptide derivs.
CC   which stimulate T-cells in persons allergic to cats.  The peptides
CC   can be used to reduce/eliminate the allergic response partic. by
CC   modificn. of lynphokine prodn. by the T-cells.  They can also be
CC   used to identify epitopes responsible for sensitivity.  The DNA can
CC   be used to detect comparable sequence in other species, and also
CC   for prodn. of modified forms of TRFP esp. showing reduced binding
CC   to IgE and thus reduced tendency to cause adverse reactions.
CC   See also AAR12119-R12123.
XX
SQ   Sequence   111 AA;


  Query Match             99.1%;  Score 547;  DB 12;  Length 111;
  Best Local Similarity   99.1%;  Pred. No. 7.8e-57;
  Matches  110;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||| ||||||||||||
Db     61 mkkiqdcyvenglisrvldglvmttissskdcmgeavqnnvedlklntlgr 111


RESULT   7
AAR36541
ID   AAR36541 standard; Protein; 109 AA.
XX
AC   AAR36541;
XX
DT   12-AUG-1993  (first entry)
XX
DE   TRFP chain 2 (with Leader).
XX
KW   Human T cell reactive feline protein; TRFP; leader A; leader B;
KW   epitope.
XX
OS   Felis.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..17
FT                   /label= leader_peptide
XX
PN   WO9308280-A.
XX
PD   29-APR-1993.
XX
PF   16-OCT-1992;   92WO-US08694.
XX
PR   16-OCT-1991;   91US-0777859.
PR   13-DEC-1991;   91US-0807529.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Bond JF,  Garman RD,  Kuo M,  Morgenstern JP,  Morville M;
PI   Rogers BL;
XX
DR   WPI; 1993-152473/18.
DR   N-PSDB; AAQ41558.
XX
PT   Recombitope peptide having T-cell stimulating activity - for the
```

PT diagnosis and treatment of sensitivity to protein allergens,
PT auto:antigens and protein antigens
XX
PS Disclosure; Fig 2; 73pp; English.
XX
CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
CC coli and purified. T cell epitope studies using overlapping peptide
CC regions derived from the TRFP amino acids sequence were used to
CC identify multiple T cell epitopes in each chain of TRFP.
XX
SQ Sequence 109 AA;

Query Match 98.0%; Score 541; DB 14; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.9e-56;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 mrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertamk 60

Qy 63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 kiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 109

RESULT 8
AAR41985
ID AAR41985 standard; Protein; 109 AA.
XX
AC AAR41985;
XX
DT 21-APR-1994 (first entry)
XX
DE Human T cell reactive feline protein chain 2.
XX
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..17
FT /note= "Signal peptide"
FT Protein 18..109
FT /note= "Mature protein"
XX
PN WO9319178-A.
XX
PD 30-SEP-1993.
XX
PF 25-MAR-1993; 93WO-US02462.
XX
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Briner TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;
PI Morville M;
XX
DR WPI; 1993-320744/40.
DR N-PSDB; AAQ49535.
XX
PT New peptide(s) for inducing tolerance - comprise one or more
PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
XX
PS Disclosure; Fig 2; 107pp; English.
XX
CC This sequence represents chain 2 of human T cell reactive feline
CC protein (TRFP). Peptides derived from TRFP may be used in a
CC therapeutic composition which is useful in treating diseases which
CC involve an immune response to a protein antigen. This composition
CC may be used to induce tolerance in a mammal to Dermatophagoides,
CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
CC Quercus, Olea, Artemesia, Plantago, Parietaria, Canis, Blattella,
CC Apis, Periplaneta and to autoantigens in humans.
XX
SQ Sequence 109 AA;

Query Match 98.0%; Score 541; DB 14; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.9e-56;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 mrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertamk 60

Qy 63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 kiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 109

RESULT 9
AAY25593
ID AAY25593 standard; protein; 109 AA.
XX
AC AAY25593;
XX
DT 30-SEP-1999 (first entry)
XX
DE Felis sp. allergen 1082946 Fel dI chain 2 precursor protein fragment.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;
KW chiromidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Felis sp.
XX
PN WO9934826-A1.
XX
PD 15-JUL-1999.
XX
PF 11-JAN-1999; 99WO-GB00080.
XX
PR 21-SEP-1998; 98GB-0020474.
PR 09-JAN-1998; 98GB-0000445.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Kay AB, Larche M;
XX
DR WPI; 1999-458255/38.
XX
PT Desensitizing patients to polypeptide allergens
XX
PS Example 6; Page 53; 117pp; English.
XX
CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitising patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chiromidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,

```
CC   non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC   Tenibrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC   sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC   produce immunological vaccines which may be used to prevent and/or treat
CC   conditions involving hypersensitivity to allergens. This sequence
CC   represents the cat allergen 1082946 Fel dI chain 2 precursor.
XX
SQ   Sequence   109 AA;


  Query Match             98.0%;  Score 541;  DB 20;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-56;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnatepertamk 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 kiqdcyvenglisrvldglvmttissskdcmgeavqntvedlklntlgr 109


RESULT  10
AAR27370
ID   AAR27370 standard; protein; 110 AA.
XX
AC   AAR27370;
XX
DT   20-MAY-1998  (first entry)
XX
DE   TRFP Chain #2 with C2 leader sequence.
XX
KW   T cell reactive feline protein; cat allergy; allergic; IgE;
KW   desensitizing.
XX
OS   Felis domesticus.
XX
FH   Key             Location/Qualifiers
FT   peptide         1..19
FT                   /label= C2 leader
FT   protein         19..110
FT                   /label= TRFP chain #2
XX
PN   WO9215613-A.
XX
PD   17-SEP-1992.
XX
PF   20-FEB-1992;   92WO-US01344.
XX
PR   28-FEB-1991;   91US-0662193.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Bond J, Kuo M;
XX
DR   WPI; 1992-331670/40.
XX
PT   Modified human T-cell reactive feline protein - stimulates T-cell
PT   in individuals allergic to cats and shows reduced
PT   histamine-releasing properties
XX
PS   Claim 1; Fig 1; 35pp; English.
XX
CC   This sequence represents a modified human T-cell reactive feline
CC   protein which stimulates T-cells from an individual who is allergic
CC   to cats, but which interacts with human IgE to a lesser extent than
CC   does affinity purified TRFP. The protein is modified by treating
CC   with either a mild alkali (pH 12.5-13.5 , KOH, NaOH, LiOH or tertiary
CC   amines) or an enzyme which removes O-linked groups (carbohydrate
CC   moieties). It is useful in desensitising people who are allergic to cats.
XX
SQ   Sequence   110 AA;


  Query Match             92.8%;  Score 512.5;  DB 13;  Length 110;
  Best Local Similarity   94.6%;  Pred. No. 8.9e-53;
  Matches  105;  Conservative    1;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||  | :  ||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvmiai-neydcmgeavqntvedlklntlgr 110


RESULT  11
AAB28934
ID   AAB28934 standard; Protein; 109 AA.
XX
AC   AAB28934;
XX
DT   29-JAN-2001  (first entry)
XX
DE   T cell reactive feline protein chain 2 short form.
XX
KW   Cat; allergy; human T cell reactive feline protein; hTRFP;
KW   immunotherapy.
XX
OS   Felis sp.
XX
PN   US6120769-A.
XX
PD   19-SEP-2000.
XX
PF   28-APR-1995;   95US-0431184.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Gefter ML,  Garman RD,  Greenstein JL,  Bond JF;
XX
DR   WPI; 2000-601477/57.
DR   N-PSDB; AAC60104.
XX
PT   Detecting, preventing and treating sensitivity to cat protein allergen
PT   comprises combining a biological sample with a human T cell reactive
PT   feline protein and determining the extent of binding that occurs  -
XX
PS   Claim ; Figure 4; 106pp; English.
XX
CC   The present invention relates to the detection of sensitivity to a cat
CC   protein allergen by combining a blood sample from a subject with a
CC   peptide of human T cell reactive feline protein (hTRFP). This method
CC   and the hTRFP peptides are useful for diagnosing, preventing and
CC   treating cat allergies by reducing or abolishing an individual's
CC   allergic response to a cat allergen. DNA encoding the TRFP may be
CC   used as probes to locate equivalent sequences present in other species.
CC   These may further be used to study the mechanism of immunotherapy of
CC   cat allergy, and to design modified derivatives, analogues or
CC   functional equivalents useful in immunotherapy. The present
CC   sequence was used in the invention.
XX
SQ   Sequence   109 AA;


  Query Match             91.7%;  Score 506;  DB 21;  Length 109;
```

```
  Best Local Similarity   93.7%;  Pred. No. 5.1e-52;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvm--iaineycmgeavqntvedlklntlgr 109


RESULT  12
AAY87674
ID   AAY87674 standard; Protein; 109 AA.
XX
AC   AAY87674;
XX
DT   22-AUG-2000  (first entry)
XX
DE   Feline human TRFP chain 2 short form protein.
XX
KW   T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
KW   antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX
OS   Felis sp.
XX
PN   US6048962-A.
XX
PD   11-APR-2000.
XX
PF   27-APR-1995;   95US-0430014.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Kuo M,  Rogers BL,  Gefter ML,  Morgenstern JP,  Brauer AW;
PI   Greenstein JL,  Griffith IJ,  Garman RD;
XX
DR   WPI; 2000-316905/27.
DR   N-PSDB; AAA12245.
XX
PT   New human T cell reactive feline protein useful for reducing or
PT   abolishing individual's allergic response to cat allergen comprising
PT   two different covalently linked peptide chains  -
XX
PS   Claim 2; Column 77-78; 106pp; English.
XX
CC   This invention describes a novel naturally occurring cat protein allergen
CC   (I), human T cell reactive feline protein (TRFP), comprising two
CC   different covalently linked peptide chains with a molecular weight of 20
CC   kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
CC   under reducing conditions. The products of the invention have
CC   antiallergic activity and act as human T cell stimulators. TRFP is useful
CC   for reducing or preventing the adverse effects of cat allergens on cat
CC   allergic individuals and in ex vivo diagnostic tests to determine which
CC   peptides cause sensitivity so as to selectively use them to desensitize
CC   a cat sensitive individual. Purified TRFP is also useful for studying
CC   the mechanism of immunotherapy of cat allergy and to design modified
CC   derivatives, analogs or functional equivalents that are more useful in
CC   immunotherapy against cat allergy. DNA sequences encoding TRFP are
CC   useful as probes to locate equivalent sequences present in other species
CC   (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC   and/or therapeutics. Fully defined and characterized TRFP provides
CC   complete and a very simple desensitization therapy. This sequence
CC   represents a human T cell reactive feline protein (also known as Fel d I)
CC   chain 2, short form which is described in the method of the invention.
XX
SQ   Sequence   109 AA;


  Query Match             91.7%;  Score 506;  DB 21;  Length 109;
  Best Local Similarity   93.7%;  Pred. No. 5.1e-52;
  Matches  104;  Conservative    3;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||  |: :: ||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldglvm--iaineycmgeavqntvedlklntlgr 109


RESULT  13
AAY90104
ID   AAY90104 standard; Protein; 109 AA.
XX
AC   AAY90104;
XX
DT   13-JUL-2000  (first entry)
XX
DE   Cat TRFP chain 2 short form protein sequence.
XX
KW   Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW   house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW   diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS   Felis domesticus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..111
FT                   /note= "mature TRFP chain 2 short form"
XX
PN   US6025162-A.
XX
PD   15-FEB-2000.
XX
PF   28-APR-1995;   95US-0430944.
XX
PR   02-SEP-1994;   94US-0300928.
PR   03-NOV-1989;   89US-0431565.
PR   28-FEB-1991;   91US-0662276.
PR   13-DEC-1991;   91US-0807529.
PR   25-MAR-1992;   92US-0857311.
PR   15-MAY-1992;   92US-0884718.
PR   15-JAN-1993;   93US-0006116.
XX
PA   (IMMU-) IMMULOGIC PHARM CORP.
XX
PI   Morgenstern JP,  Griffith IJ,  Rogers BL;
XX
DR   WPI; 2000-181812/16.
DR   N-PSDB; AAA07438.
XX
PT   New human T cell reactive feline protein, useful for desensitizing cat
PT   allergic individuals to cat allergens -
XX
PS   Claim 1; Fig 4; 108pp; English.
XX
CC   This sequence is a peptide chain of the human T cell reactive feline
CC   protein (TRFP) of the invention. The protein is a cat protein allergen,
CC   and was isolated from a vacuum bag extract obtained by affinity
CC   purification of house dust collected from several homes with cats. TRFP
CC   is composed of two covalently linked peptide chains, and is also referred
CC   to as Fel d I. TRFP and its peptides are useful for reducing or
CC   preventing the adverse effects that exposure to cat allergens normally
```

CC has on cat allergic individuals (i.e. to desensitise individuals to cat
CC allergens or block the effect of the allergens). TRFP is also used in
CC methods of diagnosing sensitivity to Felis domesticus in an individual.
CC DNA sequences encoding TRFP can be used as probes to locate equivalent
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC dog, that may be useful in a diagnostic and/or therapeutic applications.
XX
SQ Sequence 109 AA;

Query Match 91.7%; Score 506; DB 21; Length 109;
Best Local Similarity 93.7%; Pred. No. 5.1e-52;
Matches 104; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnelldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
Db 61 mkkiqdcyvenglisrvldglvm--iaineycmgeavqntvedlklntlgr 109

RESULT 14
AAY51471
ID AAY51471 standard; Protein; 109 AA.
XX
AC AAY51471;
XX
DT 22-MAY-2000 (first entry)
XX
DE Human TRFP chain 2 (short form) protein fragment.
XX
KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;
KW down regulation; immune response; allergen; immunoglobulin E;
KW sensitivity; cat protein allergen; human; chain 2.
XX
OS Homo sapiens.
XX
PN US6019972-A.
XX
PD 01-FEB-2000.
XX
PF 02-SEP-1994; 94US-0300928.
XX
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gefter ML;
XX
DR WPI; 2000-146862/13.
DR N-PSDB; AAZ88618.
XX
PT Peptides of human T cell reactive feline protein for treating
PT sensitivity to cat protein allergens comprise at least one T cell
PT epitope recognized by a T cell receptor specific for the human T cell
PT reactive feline protein -
XX
PS Claim 1; Column 79-80; 105pp; English.
XX
CC This invention describes a novel peptide (I) of human T cell reactive
CC feline protein (hTRFP) having at least one T cell epitope recognized
CC by a T cell receptor specific for the human T cell reactive feline
CC protein, the peptide consisting of at least 7-30 amino acids, and having
CC an amino acid sequence derived from an amino acid sequence comprising 94,
CC 96, 97, 109, or 111 residues, given in the specification. The peptides
CC down regulate the immune response to the allergen. The peptides have
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The
CC peptide (I) is useful in compositions for treating sensitivity to a cat
CC protein allergen in a subject. This sequence represents the human TRFP
CC chain 2 (short form).
XX
SQ Sequence 109 AA;

Query Match 91.7%; Score 506; DB 21; Length 109;
Best Local Similarity 93.7%; Pred. No. 5.1e-52;
Matches 104; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

Qy 1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
Db 1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnelldlsltkvnateperta 60

Qy 61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
Db 61 mkkiqdcyvenglisrvldglvm--iaineycmgeavqntvedlklntlgr 109

RESULT 15
AAR12122
ID AAR12122 standard; Protein; 109 AA.
XX
AC AAR12122;
XX
DT 26-JUL-1991 (first entry)
XX
DE TRFP I chain 2 - short form.
XX
KW Human T cell reactive feline protein; cat allergens.
XX
OS Felis catus.
XX
FH Key Location/Qualifiers
FT Peptide 3..19
FT /label= Leader sequence
FT Protein 20..111
FT /label= TRFP I short form
XX
PN WO9106571-A.
XX
PD 16-MAY-1991.
XX
PF 02-NOV-1990; 90WO-US06548.
XX
PR 03-NOV-1989; 89US-0431565.
XX
PA (IMMU-) IMMULOGIC PHARM COR.
XX
PI Gefter ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
PI Brauer AW;
XX
DR WPI; 1991-164136/22.
DR N-PSDB; AAQ11839.
XX
PT New pure covalently linked human T cell reactive feline protein -
PT and modified peptide(s), used to reduce effects of cat allergens
PT and to diagnose sensitivity to allergens.
XX
PS Claim 2; Fig 4; 70pp; English.
XX
CC Poly-A mRNA from cat parotid and mandibular glands was used to
CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC clones were then used to screen a cat genomic library. Chain 1
CC exists in two forms having different leader sequences (A and B).
CC The sequence can be used to express the protein and peptide derivs.
CC which stimulate T-cells in persons allergic to cats. The peptides
CC can be used to reduce/eliminate the allergic response partic. by
CC modificn. of lynphokine prodn. by the T-cells. They can also be
CC used to identify epitopes responsible for sensitivity. The DNA can
CC be used to detect comparable sequence in other species, and also

CC for prodn. of modified forms of TRFP esp. showing reduced binding
CC to IgE and thus reduced tendency to cause adverse reactions.
CC See also AAR12119-R12123.
XX
SQ Sequence 109 AA;

Query Match 90.2%; Score 498; DB 12; Length 109;
Best Local Similarity 92.8%; Pred. No. 4.5e-51;
Matches 103; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

```
Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 dtmrgallvlallvtqalgvkmaetcpifydvffavangnellldlsltkvnateperta 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||| ||| |: :: ||||||||||||||||||||
Db      61 mkkiqdcyvenglisrvldelvm--iaineycmgeavqntvedlklntlgr 109
```

Search completed: May 7, 2002, 12:13:28
Job time: 200 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:16:04 ; Search time 12.5 Seconds
(without alignments)
199.829 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 111
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	111	100.0	111	1	US-07-662-193-4	Sequence 4, Appli
2	111	100.0	111	3	US-08-300-928C-6	Sequence 6, Appli
3	111	100.0	111	3	US-08-430-944D-6	Sequence 6, Appli
4	111	100.0	111	3	US-08-430-014-6	Sequence 6, Appli
5	111	100.0	111	3	US-08-431-184-6	Sequence 6, Appli
6	109	98.2	109	1	US-07-807-529A-6	Sequence 6, Appli
7	109	98.2	109	5	PCT-US93-02462-6	Sequence 6, Appli
8	92	82.9	92	3	US-08-300-928C-13	Sequence 13, Appl
9	92	82.9	92	3	US-08-430-944D-13	Sequence 13, Appl
10	92	82.9	92	3	US-08-430-014-13	Sequence 13, Appl
11	92	82.9	92	3	US-08-431-184-13	Sequence 13, Appl
12	83	74.8	109	1	US-07-662-193-5	Sequence 5, Appli
13	83	74.8	109	3	US-08-300-928C-8	Sequence 8, Appli
14	83	74.8	109	3	US-08-430-944D-8	Sequence 8, Appli
15	83	74.8	109	3	US-08-430-014-8	Sequence 8, Appli
16	83	74.8	109	3	US-08-431-184-8	Sequence 8, Appli
17	81	73.0	97	3	US-08-300-928C-10	Sequence 10, Appl
18	81	73.0	97	3	US-08-430-944D-10	Sequence 10, Appl
19	81	73.0	97	3	US-08-430-014-10	Sequence 10, Appl
20	81	73.0	97	3	US-08-431-184-10	Sequence 10, Appl
21	64	57.7	90	3	US-08-300-928C-14	Sequence 14, Appl
22	64	57.7	90	3	US-08-430-944D-14	Sequence 14, Appl
23	64	57.7	90	3	US-08-430-014-14	Sequence 14, Appl
24	64	57.7	90	3	US-08-431-184-14	Sequence 14, Appl
25	37	33.3	51	1	US-07-662-193-6	Sequence 6, Appli
26	37	33.3	51	3	US-08-300-928C-15	Sequence 15, Appl
27	37	33.3	51	3	US-08-430-944D-15	Sequence 15, Appl
28	37	33.3	51	3	US-08-430-014-15	Sequence 15, Appl
29	37	33.3	51	3	US-08-431-184-15	Sequence 15, Appl
30	36	32.4	83	1	US-07-662-193-7	Sequence 7, Appli
31	32	28.8	82	3	US-08-300-928C-16	Sequence 16, Appl
32	32	28.8	82	3	US-08-430-944D-16	Sequence 16, Appl
33	32	28.8	82	3	US-08-430-014-16	Sequence 16, Appl
34	32	28.8	82	3	US-08-431-184-16	Sequence 16, Appl
35	32	28.8	83	1	US-07-662-193-8	Sequence 8, Appli
36	26	23.4	26	1	US-07-807-529A-9	Sequence 9, Appli
37	26	23.4	26	3	US-08-300-928C-19	Sequence 19, Appl
38	26	23.4	26	3	US-08-430-944D-19	Sequence 19, Appl
39	26	23.4	26	3	US-08-430-014-19	Sequence 19, Appl
40	26	23.4	26	3	US-08-431-184-19	Sequence 19, Appl
41	26	23.4	26	5	PCT-US93-02462-9	Sequence 9, Appli
42	26	23.4	96	1	US-07-807-529A-39	Sequence 39, Appl
43	26	23.4	96	3	US-08-430-944D-103	Sequence 103, App
44	26	23.4	96	3	US-08-431-184-103	Sequence 103, App
45	23	20.7	27	3	US-08-300-928C-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-07-662-193-4
; Sequence 4, Application US/07662193
; Patent No. 5328991
; GENERAL INFORMATION:
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Bond, Julian
; TITLE OF INVENTION: Improved Preparation of Cat Dander
; TITLE OF INVENTION: Allergens for Immunotherapeutic Purposes and Uses Thereof
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/662,193
; FILING DATE: 19910228
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,276
; FILING DATE: 28-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: P36,207
; REFERENCE/DOCKET NUMBER: IML89-02AA/IPC-002CC/IMI-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-07-662-193-4

Query Match 100.0%; Score 111; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.1e-102;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   2
US-08-300-928C-6
; Sequence 6, Application US/08300928C
; Patent No. 6019972
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,928C
;      FILING DATE:  September 2, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,529
;      FILING DATE:  December 13, 1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  111 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-300-928C-6


  Query Match             100.0%;  Score 111;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-102;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   3
US-08-430-944D-6
; Sequence 6, Application US/08430944D
; Patent No. 6025162
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,944D
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  111 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-430-944D-6


  Query Match             100.0%;  Score 111;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-102;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   4
US-08-430-014-6
; Sequence 6, Application US/08430014
; Patent No. 6048962
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/430,014
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/300,928
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  AMY E. MANDRAGOURAS
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  111 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-430-014-6


  Query Match             100.0%;  Score 111;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-102;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   5
US-08-431-184-6
; Sequence 6, Application US/08431184
; Patent No. 6120769
;  GENERAL INFORMATION:
;     APPLICANT:  Bruce L. Rogers et al.
;     TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;     TITLE OF INVENTION:
;     NUMBER OF SEQUENCES:  103
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  28 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/431,184
;       FILING DATE:  28-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/430,014
;       FILING DATE:  27-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/300,928
;       FILING DATE:  02-SEPT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amy E. Mandragouras
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  IMI-044DV3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)742-4214
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  111 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-431-184-6


  Query Match             100.0%;  Score 111;  DB 3;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-102;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111


RESULT   6
US-07-807-529A-6
; Sequence 6, Application US/07807529A
; Patent No. 5547669
;  GENERAL INFORMATION:
;     APPLICANT:  Rogers, Bruce L.
;     APPLICANT:  Morgenstern, Jay
;     APPLICANT:  Bond, Julian F.
;     APPLICANT:  Garman, Richard D.
;     APPLICANT:  Greenstein, Julia L.
;     APPLICANT:  Kuo, Mei-chang
;     APPLICANT:  Morvile, Malcolm
;     TITLE OF INVENTION:  RECOMBITOPE PEPTIDES
;     NUMBER OF SEQUENCES:  76
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;       STREET:  One Kendall Square, Building 600
;       CITY:  Cambridge
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02139
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII TEXT
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/807,529A
;       FILING DATE:  19911213
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/662,276
;       FILING DATE:  28-FEB-1991
;       APPLICATION NUMBER:  US 07/431,565
;       FILING DATE:  03-NOV-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Channing, Stacey L.
;       REGISTRATION NUMBER:  31,095
;       REFERENCE/DOCKET NUMBER:  IPC-027/imi-015
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 494-0060
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  109 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-807-529A-6
```

```
  Query Match             98.2%;  Score 109;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-100;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109


RESULT   7
PCT-US93-02462-6
; Sequence 6, Application PC/TUS9302462
;  GENERAL INFORMATION:
;    APPLICANT:  Gefter, Malcolm L.
;    APPLICANT:  Garman, Richard D.
;    APPLICANT:  Greenstein, Julia L.
;    APPLICANT:  Kuo, Mei-chang
;    APPLICANT:  Briner, Thomas J.
;    APPLICANT:  Morville, Malcolm
;    TITLE OF INVENTION:  PEPTIDES USEFUL FOR TOLERIZATION
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02462
;      FILING DATE:  19930325
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/006,116
;      FILING DATE:  15-JAN-1993
;      APPLICATION NUMBER:  US 07/884,718
;      FILING DATE:  15-MAY-1992
;      APPLICATION NUMBER:  07/857,311
;      FILING DATE:  25-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-031PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-02462-6


  Query Match             98.2%;  Score 109;  DB 5;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-100;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109


RESULT   8
US-08-300-928C-13
; Sequence 13, Application US/08300928C
; Patent No. 6019972
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,928C
;      FILING DATE:  September 2, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,529
;      FILING DATE:  December 13, 1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-300-928C-13


  Query Match             82.9%;  Score 92;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-84;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


RESULT   9
US-08-430-944D-13
; Sequence 13, Application US/08430944D
; Patent No. 6025162
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  28 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/430,944D
;       FILING DATE:  28-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/430,014
;       FILING DATE:  27-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/300,928
;       FILING DATE:  02-SEPT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Amy E. Mandragouras
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  IMI-044DV2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)742-4214
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  92 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  internal
US-08-430-944D-13


  Query Match             82.9%;  Score 92;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-84;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


RESULT  10
US-08-430-014-13
; Sequence 13, Application US/08430014
; Patent No. 6048962
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,014
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/300,928
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-430-014-13


  Query Match             82.9%;  Score 92;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-84;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


RESULT  11
US-08-431-184-13
; Sequence 13, Application US/08431184
; Patent No. 6120769
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/431,184
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
```

```
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-431-184-13


  Query Match             82.9%;  Score 92;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-84;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 79
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMKKIQDCYVENGLISRVLD 60

Qy      80 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           ||||||||||||||||||||||||||||||||
Db      61 GLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 92


RESULT  12
US-07-662-193-5
; Sequence 5, Application US/07662193
; Patent No. 5328991
;  GENERAL INFORMATION:
;    APPLICANT:  Kuo, Mei-chang
;    APPLICANT:  Bond, Julian
;    TITLE OF INVENTION:  Improved Preparation of Cat Dander
;    TITLE OF INVENTION:  Allergens for Immunotherapeutic Purposes and Uses Thereof
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/662,193
;      FILING DATE:  19910228
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/662,276
;      FILING DATE:  28-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  P36,207
;      REFERENCE/DOCKET NUMBER:  IML89-02AA/IPC-002CC/IMI-020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-07-662-193-5


  Query Match             74.8%;  Score 83;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-75;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM 83


RESULT  13
US-08-300-928C-8
; Sequence 8, Application US/08300928C
; Patent No. 6019972
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,928C
;      FILING DATE:  September 2, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,529
;      FILING DATE:  December 13, 1991
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-300-928C-8


  Query Match             74.8%;  Score 83;  DB 3;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-75;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM 83


RESULT  14
US-08-430-944D-8
; Sequence 8, Application US/08430944D
; Patent No. 6025162
;  GENERAL INFORMATION:
;    APPLICANT:  Bruce L. Rogers et al.
```

```
;    TITLE OF INVENTION:  A HUMAN T CELL REACTIVE FELINE PROTEIN
;    TITLE OF INVENTION:
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,944D
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/430,014
;      FILING DATE:  27-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,928
;      FILING DATE:  02-SEPT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-044DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-430-944D-8


  Query Match             74.8%;  Score 83;  DB 3;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-75;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM 83


RESULT  15
US-08-430-014-8
; Sequence 8, Application US/08430014
; Patent No. 6048962
;  GENERAL INFORMATION:
;    APPLICANT:  GEFTER, Malcolm L. et al.
;    TITLE OF INVENTION:  PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
;    TITLE OF INVENTION:  PROTEIN (TRFP)
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;      STREET:  610 LINCOLN STREET
;      CITY:  WALTHAM
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02145
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,014
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/300,928
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  AMY E. MANDRAGOURAS
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  002.6US(IMI-044)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  109 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-430-014-8


  Query Match             74.8%;  Score 83;  DB 3;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-75;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 DTMRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTA 60

Qy      61 MKKIQDCYVENGLISRVLDGLVM 83
           |||||||||||||||||||||||
Db      61 MKKIQDCYVENGLISRVLDGLVM 83


Search completed: May  7, 2002, 12:17:53
Job time: 109 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:17:39 ; Search time 22.16 Seconds
(without alignments)
732.682 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 111
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	7.2	119	13	Q9YGD0	Q9ygd0 petromyzon
2	8	7.2	138	13	Q9PSW3	Q9psw3 petromyzon
3	8	7.2	138	13	Q9PSW2	Q9psw2 petromyzon
4	8	7.2	181	2	O87052	O87052 vibrio chol
5	8	7.2	266	2	Q9KW39	Q9kw39 wolbachia s
6	8	7.2	318	11	Q99NE4	Q99ne4 mus musculu
7	8	7.2	322	4	Q9Y5Y7	Q9y5y7 homo sapien
8	8	7.2	322	4	Q9UNF4	Q9unf4 homo sapien
9	8	7.2	390	2	Q9AEX7	Q9aex7 treponema h
10	8	7.2	390	2	Q9AEX5	Q9aex5 treponema h
11	8	7.2	401	2	Q9HT26	Q9ht26 pseudomonas
12	7	6.3	61	5	O61029	O61029 trypanosoma
13	7	6.3	64	5	O61042	O61042 trypanosoma
14	7	6.3	73	5	Q9XU39	Q9xu39 caenorhabdi
15	7	6.3	74	5	O61023	O61023 trypanosoma
16	7	6.3	77	2	Q9KG69	Q9kg69 bacillus ha
17	7	6.3	78	5	Q26881	Q26881 trypanosoma
18	7	6.3	87	2	Q03539	Q03539 escherichia
19	7	6.3	88	5	O61047	O61047 trypanosoma
20	7	6.3	94	5	O61030	O61030 trypanosoma
21	7	6.3	102	5	O61033	O61033 trypanosoma
22	7	6.3	103	2	Q9RRB9	Q9rrb9 deinococcus
23	7	6.3	107	5	O61050	O61050 trypanosoma
24	7	6.3	112	5	Q9VLE8	Q9vle8 drosophila
25	7	6.3	113	5	O61052	O61052 trypanosoma
26	7	6.3	115	2	Q9RD28	Q9rd28 streptomyce
27	7	6.3	115	5	O61046	O61046 trypanosoma
28	7	6.3	119	5	O61034	O61034 trypanosoma
29	7	6.3	122	5	O15774	O15774 trypanosoma
30	7	6.3	123	5	P90602	P90602 trypanosoma
31	7	6.3	123	5	O15773	O15773 trypanosoma
32	7	6.3	123	5	O61027	O61027 trypanosoma
33	7	6.3	125	5	O61025	O61025 trypanosoma
34	7	6.3	126	5	O61021	O61021 trypanosoma
35	7	6.3	126	5	O61056	O61056 trypanosoma
36	7	6.3	127	5	O61035	O61035 trypanosoma
37	7	6.3	128	5	P90603	P90603 trypanosoma
38	7	6.3	133	5	O61022	O61022 trypanosoma
39	7	6.3	136	5	O15775	O15775 trypanosoma
40	7	6.3	139	5	P90601	P90601 trypanosoma
41	7	6.3	139	5	O61037	O61037 trypanosoma
42	7	6.3	139	13	P79799	P79799 micrurus co
43	7	6.3	143	5	O15776	O15776 trypanosoma
44	7	6.3	147	5	O61043	O61043 trypanosoma
45	7	6.3	163	5	O18679	O18679 trypanosoma

ALIGNMENTS

RESULT 1
Q9YGD0
ID Q9YGD0 PRELIMINARY; PRT; 119 AA.
AC Q9YGD0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE LAMPRIN L-0.9-10.
GN L-0.9.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93123269; PubMed=7678258;
RA Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA Keeley F.W.;
RT "Characterization of lamprin, an unusual matrix protein from lamprey
RT cartilage. Implications for evolution, structure, and assembly of
RT elastin and other fibrillar proteins.";
RL J. Biol. Chem. 268:1440-1447(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Robson P., Youson J.H., Keeley F.W.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF097759; AAC97500.1; -.
DR EMBL; AF097758; AAC97500.1; JOINED.
DR EMBL; AF097757; AAC97498.1; -.
DR EMBL; AF097756; AAC97498.1; JOINED.
SQ SEQUENCE 119 AA; 11145 MW; BB608EE263DF03E0 CRC64;

Query Match 7.2%; Score 8; DB 13; Length 119;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVLALL 13
||||||||
Db 7 ALLVLALL 14

```
RESULT   2
Q9PSW3
ID   Q9PSW3      PRELIMINARY;      PRT;   138 AA.
AC   Q9PSW3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LAMPRIN L-0.8-12.
GN   L-0.8.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robson P., Youson J.H., Keeley F.W.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF097757; AAC97499.1; -.
DR   EMBL; AF097756; AAC97499.1; JOINED.
SQ   SEQUENCE   138 AA;  13094 MW;  AB2F35D0D25D7580 CRC64;

  Query Match             7.2%;  Score 8;  DB 13;  Length 138;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 ALLVLALL 13
           ||||||||
Db       7 ALLVLALL 14


RESULT   3
Q9PSW2
ID   Q9PSW2      PRELIMINARY;      PRT;   138 AA.
AC   Q9PSW2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LAMPRIN L-0.9-12.
GN   L-0.9.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robson P., Youson J.H., Keeley F.W.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF097759; AAC97501.1; -.
DR   EMBL; AF097758; AAC97501.1; JOINED.
SQ   SEQUENCE   138 AA;  13144 MW;  AB2F35D0C74924D0 CRC64;

  Query Match             7.2%;  Score 8;  DB 13;  Length 138;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 ALLVLALL 13
           ||||||||
Db       7 ALLVLALL 14


RESULT   4
O87052
ID   O87052      PRELIMINARY;      PRT;   181 AA.
AC   O87052;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Z48F PROTEIN (FRAGMENT).
GN   Z48F.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z17561;
RA   Fallarino A.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ231108; CAA13150.1; -.
DR   InterPro; IPR003764; NagA.
DR   Pfam; PF02612; NagA; 1.
FT   NON_TER       1       1
FT   NON_TER     181     181
SQ   SEQUENCE   181 AA;  19476 MW;  8527718DAC9DD199 CRC64;

  Query Match             7.2%;  Score 8;  DB 2;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 EAVQNTVE 102
           ||||||||
Db     150 EAVQNTVE 157


RESULT   5
Q9KW39
ID   Q9KW39      PRELIMINARY;      PRT;   266 AA.
AC   Q9KW39;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   COMPONENT OF TYPE IV SECRETION SYSTEM.
GN   VIRB9.
OS   Wolbachia sp. wKueYO.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Wolbachieae; Wolbachia.
OX   NCBI_TaxID=130485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WKUEYO;
RA   Masui S.;
RT   "vir genes for the type IV secretion system of Wolbachia wKueYO.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045235; BAA97440.1; -.
SQ   SEQUENCE   266 AA;  31112 MW;  23C42E77A756B5E2 CRC64;

  Query Match             7.2%;  Score 8;  DB 2;  Length 266;
  Best Local Similarity   100.0%;  Pred. No. 6.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALLV 14
           ||||||||
Db       7 LLVLALLV 14
```

```
RESULT   6
Q99NE4
ID   Q99NE4      PRELIMINARY;      PRT;   318 AA.
AC   Q99NE4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYALURONAN RECEPTOR PRECURSOR.
GN   LYVE-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=DIGESTIVE TRACT;
RX   MEDLINE=99156989; PubMed=10037799;
RA   Banerji S.;
RT   "LYVE-1, a new homologue of the CD44 glycoprotein is a lymph-specific
RT   receptor for hyaluronan. LYVE-1, a new homologue of the CD44
RT   glycoprotein is a lymph-specific receptor for hyaluronan.";
RL   J. Cell Biol. 144:789-801(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=DIGESTIVE TRACT;
RA   Prevo R., Banerji S., Ferguson D.J.P., Clasper S., Jackson D.G.;
RT   "Mouse LYVE-1 is an endocytic receptor for hyaluronan in lymphatic
RT   endothelium.";
RL   J. Biol. Chem. 0:0-0(0).
DR   EMBL; AJ311501; CAC33082.1; -.
KW   Signal; Receptor.
FT   SIGNAL        1     23
FT   CHAIN        24    318       LYVE-1.
SQ   SEQUENCE   318 AA;  34641 MW;  1248974A16113330 CRC64;


  Query Match             7.2%;  Score 8;  DB 11;  Length 318;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 ALLVLALL 13
          ||||||||
Db    235 ALLVLALL 242


RESULT   7
Q9Y5Y7
ID   Q9Y5Y7      PRELIMINARY;      PRT;   322 AA.
AC   Q9Y5Y7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYMPHATIC ENDOTHELIUM-SPECIFIC HYALURONAN RECEPTOR LYVE-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99156989; PubMed=10037799;
RA   Banerji S., Ni J., Wang S.X., Clasper S., Su J., Tammi R., Jones M.,
RA   Jackson D.G.;
RT   "LYVE-1, a new homologue of the CD44 glycoprotein, is a lymph-specific
RT   receptor for hyaluronan.";
RL   J. Cell Biol. 144:789-801(1999).
DR   EMBL; AF118108; AAD42764.1; -.
DR   HSSP; P98066; 1TSG.
DR   InterPro; IPR000538; Link.
DR   Pfam; PF00193; Xlink; 1.
DR   ProDom; PD000918; Link; 1.
DR   SMART; SM00445; LINK; 1.
KW   Receptor.
SQ   SEQUENCE   322 AA;  35238 MW;  0EBEA56729CEFFF7 CRC64;


  Query Match             7.2%;  Score 8;  DB 4;  Length 322;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 ALLVLALL 13
          ||||||||
Db    239 ALLVLALL 246


RESULT   8
Q9UNF4
ID   Q9UNF4      PRELIMINARY;      PRT;   322 AA.
AC   Q9UNF4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYALURONIC ACID RECEPTOR.
GN   HAR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Winkelmann J.C., Basu S., Ozdemir E., Blough R.I.;
RT   "HAR: a novel homolog of CD44 and putative hyaluronic acid receptor
RT   encoded by a gene on human chromosome 11p15.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF127670; AAD49220.2; -.
DR   HSSP; P98066; 1TSG.
DR   InterPro; IPR000538; Link.
DR   Pfam; PF00193; Xlink; 1.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 1.
DR   SMART; SM00445; LINK; 1.
KW   Receptor.
SQ   SEQUENCE   322 AA;  35213 MW;  8B4D6D623F52D559 CRC64;


  Query Match             7.2%;  Score 8;  DB 4;  Length 322;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 ALLVLALL 13
          ||||||||
Db    239 ALLVLALL 246


RESULT   9
Q9AEX7
ID   Q9AEX7      PRELIMINARY;      PRT;   390 AA.
AC   Q9AEX7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VARIABLE SURFACE PROTEIN VSPF.
GN   VSPF.
OS   Treponema hyodysenteriae (Serpulina hyodysenteriae).
OC   Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
OX   NCBI_TaxID=159;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   McCaman M., Gabe J.;
RT   "Serpulina hyodysenteriae (B204) contains eight gene copies related to
RT   a 39 kDa surface protein.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY027775; AAK14801.1; -.
SQ   SEQUENCE   390 AA;  42836 MW;  9B20084CE61A2B9E CRC64;


  Query Match             7.2%;  Score 8;  DB 2;  Length 390;
```

```
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      94 GEAVQNTV 101
           ||||||||
Db     247 GEAVQNTV 254


RESULT  10
Q9AEX5
ID   Q9AEX5       PRELIMINARY;      PRT;   390 AA.
AC   Q9AEX5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VARIABLE SURFACE PROTEIN VSPH.
GN   VSPH.
OS   Treponema hyodysenteriae (Serpulina hyodysenteriae).
OC   Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
OX   NCBI_TaxID=159;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   McCaman M., Gabe J.;
RT   "Serpulina hyodysenteriae (B204) contains eight gene copies related to
RT   a 39 kDa surface protein.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY027775; AAK14803.1; -.
SQ   SEQUENCE   390 AA;  42816 MW;  C3B63F5C9BC36E08 CRC64;


  Query Match             7.2%;  Score 8;  DB 2;  Length 390;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      94 GEAVQNTV 101
           ||||||||
Db     245 GEAVQNTV 252


RESULT  11
Q9HT26
ID   Q9HT26       PRELIMINARY;      PRT;   401 AA.
AC   Q9HT26;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROBABLE MFS TRANSPORTER.
GN   PA5548.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004967; AAG08933.1; -.
DR   InterPro; IPR001958; TCR_TetA.
DR   InterPro; IPR001411; TCR_TetB.
DR   PRINTS; PR01035; TCRTETA.
DR   PRINTS; PR01036; TCRTETB.
KW   Complete proteome.
SQ   SEQUENCE   401 AA;  42247 MW;  D8D9EEABDABD509E CRC64;


  Query Match             7.2%;  Score 8;  DB 2;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 9.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALLV 14
            ||||||||
Db     281 LLVLALLV 288


RESULT  12
O61029
ID   O61029       PRELIMINARY;      PRT;    61 AA.
AC   O61029;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   MUCIN-LIKE PROTEIN (FRAGMENT).
GN   EMUCE-21E22BIS.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CL-BRENNER;
RA   Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.C.;
RL   J. Biol. Chem. 0:0-0(1998).
DR   EMBL; AF036418; AAC14228.1; -.
FT   NON_TER      61      61
SQ   SEQUENCE   61 AA;  6192 MW;  95ABC7D84C05B587 CRC64;


  Query Match             6.3%;  Score 7;  DB 5;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 ALLVLAL 12
            |||||||
Db      10 ALLVLAL 16


RESULT  13
O61042
ID   O61042       PRELIMINARY;      PRT;    64 AA.
AC   O61042;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   MUCIN-LIKE PROTEIN (FRAGMENT).
GN   EMUCE-13.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CL-BRENNER;
RA   Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.C.;
RL   J. Biol. Chem. 0:0-0(1998).
DR   EMBL; AF036439; AAC14242.1; -.
FT   NON_TER      64      64
SQ   SEQUENCE   64 AA;  6673 MW;  A2B55BB7D05608FF CRC64;


  Query Match             6.3%;  Score 7;  DB 5;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 ALLVLAL 12
            |||||||
Db      14 ALLVLAL 20


RESULT  14
```

```
Q9XU39
ID   Q9XU39      PRELIMINARY;      PRT;    73 AA.
AC   Q9XU39;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   F14F8.8 PROTEIN.
GN   F14F8.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lloyd C.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z92782; CAB07183.1; -.
SQ   SEQUENCE   73 AA;  7210 MW;  2F5F985045CB141D CRC64;


  Query Match             6.3%;  Score 7;  DB 5;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 LLVLALL 13
          |||||||
Db      5 LLVLALL 11


RESULT  15
O61023
ID   O61023      PRELIMINARY;      PRT;    74 AA.
AC   O61023;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   MUCIN-LIKE PROTEIN (FRAGMENT).
GN   EMUCE-4.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CL-BRENNER;
RA   Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.C.;
RL   J. Biol. Chem. 0:0-0(1998).
DR   EMBL; AF036411; AAC14222.1; -.
FT   NON_TER      74      74
SQ   SEQUENCE   74 AA;  7743 MW;  734CC37663E21401 CRC64;


  Query Match             6.3%;  Score 7;  DB 5;  Length 74;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 ALLVLAL 12
          |||||||
Db     14 ALLVLAL 20


Search completed: May  7, 2002, 12:20:52
Job time: 193 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:17:59 ; Search time 11.68 Seconds
(without alignments)
348.441 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 111
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	109	98.2	109	1	FEL2_FELCA	P30440 felis silve
2	8	7.2	139	1	LAMP_PETMA	P33575 petromyzon
3	8	7.2	299	1	YCXC_CYAPA	P31606 cyanophora
4	8	7.2	367	1	Y4HA_RHISN	P55471 rhizobium s
5	8	7.2	378	1	NAGA_VIBCH	O32445 vibrio chol
6	8	7.2	1106	1	ITA7_RAT	Q63258 rattus norv
7	8	7.2	1179	1	ITA7_MOUSE	Q61738 mus musculu
8	8	7.2	1181	1	ITA7_HUMAN	Q13683 homo sapien
9	7	6.3	141	1	YEDD_SALTY	Q06399 salmonella
10	7	6.3	146	1	LYC2_PIG	P12068 sus scrofa
11	7	6.3	148	1	LYC3_PIG	P12069 sus scrofa
12	7	6.3	250	1	BACR_HALS4	O93740 halobacteri
13	7	6.3	322	1	QOX2_BACSU	P34957 bacillus su
14	7	6.3	332	1	C1B2_CAVPO	Q9qzz1 cavia porce
15	7	6.3	337	1	YE71_HAEIN	Q57130 haemophilus
16	7	6.3	397	1	UL50_HCMVA	P16791 human cytom
17	7	6.3	400	1	ARGD_MYCTU	P94990 mycobacteri
18	7	6.3	420	1	HISZ_SYNP7	Q55267 synechococc
19	7	6.3	424	1	ODB2_BACSU	P37942 bacillus su
20	7	6.3	434	1	GSA_CERSY	O74038 cenarchaeum
21	7	6.3	531	1	RO60_DEIRA	Q9ruw8 deinococcus
22	7	6.3	574	1	ILVI_ECOLI	P00893 escherichia
23	7	6.3	707	1	MM09_RABIT	P41246 oryctolagus
24	7	6.3	712	1	MM09_BOVIN	P52176 bos taurus
25	7	6.3	729	1	CUL6_CAEEL	Q21346 caenorhabdi
26	7	6.3	899	1	PR06_YEAST	P19735 saccharomyc
27	7	6.3	1189	1	ITAH_HUMAN	Q9ukx5 homo sapien
28	7	6.3	1345	1	VCAP_HSV6U	P17887 human herpe
29	7	6.3	3412	1	POLG_TBEVS	P07720 t genome po
30	7	6.3	3414	1	POLG_TBEVH	Q01299 t genome po
31	7	6.3	3414	1	POLG_TBEVW	P14336 t genome po
32	6	5.4	80	1	ITF_HUMAN	Q07654 homo sapien
33	6	5.4	100	1	SECG_AQUAE	O66505 aquifex aeo
34	6	5.4	101	1	VNST_BUNYW	P16494 bunyamwera
35	6	5.4	101	1	VNST_MAGV	P16605 maguari vir
36	6	5.4	106	1	KACA_RAT	P01836 rattus norv
37	6	5.4	108	1	INS_BRARE	O73727 brachydanio
38	6	5.4	114	1	DCHS_LACBU	P04193 lactobacill
39	6	5.4	115	1	MERT_SHEPU	Q54462 shewanella
40	6	5.4	120	1	CHH4_PENMO	O97386 penaeus mon
41	6	5.4	123	1	YBAV_ECOLI	P77415 escherichia
42	6	5.4	125	1	RNP_SPAEH	P16414 spalax leuc
43	6	5.4	127	1	CHMU_BACSU	P19080 bacillus su
44	6	5.4	128	1	Y070_TREPA	O83109 treponema p
45	6	5.4	144	1	VE6_COPV	Q89808 canine oral

ALIGNMENTS

RESULT 1
FEL2_FELCA
ID FEL2_FELCA STANDARD; PRT; 109 AA.
AC P30440;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 2 PRECURSOR (ALLERGEN FEL D 1-B)
DE (FEL D I-B) (ALLERGEN CAT-1) (AG4) (FDI).
GN CH2.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-100.
RX MEDLINE=92052157; PubMed=1946388;
RA Morgenstern J.P., Griffith I.J., Brauer A.W., Rogers B.L.,
RA Bond J.F., Chapman M.D., Kuo M.-C.;
RT "Amino acid sequence of Fel dI, the major allergen of the domestic
RT cat: protein sequence analysis and cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:9690-9694(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92241678; PubMed=1572548;
RA Griffith I.J., Craig S., Pollock J., Yu X.-B., Morgenstern J.P.,
RA Rogers B.L.;
RT "Expression and genomic structure of the genes encoding FdI, the
RT major allergen from the domestic cat.";
RL Gene 113:263-268(1992).
RN [3]
RP SEQUENCE OF 18-37, AND CHARACTERIZATION.
RX MEDLINE=91287714; PubMed=1712068;
RA Duffort O.A., Carreira J., Nitti G., Polo F., Lombardero M.;
RT "Studies on the biochemical structure of the major cat allergen Felis
RT domesticus I.";
RL Mol. Immunol. 28:301-309(1991).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=84265679; PubMed=6747135;
RA Leiterman K., Ohman J.L. Jr.;
RT "Cat allergen 1: Biochemical, antigenic, and allergenic properties.";
RL J. Allergy Clin. Immunol. 74:147-153(1984).
CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
CC -!- ALTERNATIVE PRODUCTS: THE LONG (CH2L) AND THE SHORT (CH2S) FORMS
CC MAY ARISE BY ALTERNATIVE SPLICING OR MAY REPRESENT DIFFERENT
CC ALLELES OF THE CH2 GENE. THE SEQUENCE SHOWN HERE IS THAT OF THE
CC LONG FORM (CH2L).
CC -!- TISSUE SPECIFICITY: THE LONG FORM IS PREFERENTIALLY EXPRESSED IN
CC THE SALIVARY GLAND, WHILE THE SHORT FORM IS PREFERENTIALLY
CC EXPRESSED IN THE SKIN.
CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
CC ----------

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M77341; AAC41616.1; -.
DR   EMBL; X62478; CAA44345.1; -.
DR   PIR; JC1127; JC1127.
DR   PIR; JC1145; JC1145.
KW   Allergen; Glycoprotein; Signal; Polymorphism; Alternative splicing.
FT   SIGNAL       1     17
FT   CHAIN       18    109       MAJOR ALLERGEN I POLYPEPTIDE CHAIN 2.
FT   CARBOHYD    50     50       N-LINKED (GLCNAC. . .).
FT   VARSPLIC    82     82       T -> PSTNIAWVKQFRTP (IN ISOFORM CH2ST).
FT   VARSPLIC    83    109       MISSING (IN ISOFORM CH2ST).
FT   VARIANT     72     72       I -> L (IN CH2LV).
FT   VARIANT     72     72       I -> V (IN CH2SV).
FT   VARIANT     74     75       RV -> KF (IN CH2SV).
FT   VARIANT     82     83       TT -> IA (IN CH2S AND CH2SV).
FT   VARIANT     85     85       S -> N (IN CH2S AND CH2SV).
FT   VARIANT     86     86       S -> E (IN CH2S).
FT   VARIANT     86     86       S -> V (IN CH2SV).
FT   VARIANT     87     88       MISSING (IN CH2S AND CH2SV).
FT   VARIANT     89     89       D -> Y (IN CH2S).
FT   VARIANT     96     96       Q -> E (IN CH2SV).
FT   VARIANT    105    105       N -> K (IN CH2SV).
FT   CONFLICT    24     24       C -> F (IN REF. 3).
FT   CONFLICT    32     32       F -> T (IN REF. 3).
SQ   SEQUENCE   109 AA;  11854 MW;  857FB9CD76036CB9 CRC64;


  Query Match             98.2%;  Score 109;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 1e-100;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109


RESULT   2
LAMP_PETMA
ID   LAMP_PETMA     STANDARD;      PRT;   139 AA.
AC   P33575; P33576;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   LAMPRIN 0.9 PRECURSOR (CARTILAGE MATRIX PROTEIN).
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-44.
RC   TISSUE=Cartilage;
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
CC   -!- FUNCTION: SELF-AGGREGATING PROTEIN THAT MAKES PART OF THE SOLUBLE
CC       FORM OF LAMPRIN.
CC   -!- SUBUNIT: THE POLYMERIC LAMPRIN CHAINS SELF-AGGREGATE TO FORM
CC       FIBERS AND HAVE SECONDARY STRUCTURES PARTICULARLY RICH IN BETA-
CC       SHEETS AND IN BETA-TURNS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORM; 0.9-12 (SHOWN HERE) AND 0.9-10;
CC       SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L05925; AAA49269.1; -.
DR   EMBL; L05924; AAA49268.1; -.
DR   PIR; C45051; C45051.
DR   PIR; A45051; A45051.
KW   Cartilage; Repeat; Connective tissue; Signal; Alternative splicing.
FT   SIGNAL       1     19
FT   CHAIN       20    139       LAMPRIN 0.9.
FT   DOMAIN      42    110       8 X 5 AA APPROXIMATE REPEATS.
FT   REPEAT      42     46       1.
FT   REPEAT      47     51       2.
FT   REPEAT      52     56       3.
FT   REPEAT      57     61       4.
FT   REPEAT      62     66       5.
FT   REPEAT      67     71       6.
FT   REPEAT      92     96       7.
FT   REPEAT     106    110       8.
FT   VARSPLIC    86    104       MISSING (IN ISOFORM 0.9-10).
SQ   SEQUENCE   139 AA;  13257 MW;  E248ABB7A6CEC7C3 CRC64;


  Query Match             7.2%;  Score 8;  DB 1;  Length 139;
  Best Local Similarity   100.0%;  Pred. No. 0.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 ALLVLALL 13
             ||||||||
Db       7 ALLVLALL 14


RESULT   3
YCXC_CYAPA
ID   YCXC_CYAPA     STANDARD;      PRT;   299 AA.
AC   P31606;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 32.8 KDA PROTEIN IN YCF23-APCF INTERGENIC REGION
DE   (ORF299).
OS   Cyanophora paradoxa.
OG   Cyanelle.
OC   Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX   NCBI_TaxID=2762;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LB555 / PRINGSHEIM;
RA   Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA   Bryant D.A.;
RT   "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL   Plant Mol. Biol. Rep. 13:327-332(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LB555 / PRINGSHEIM;
RA   Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA   Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA   Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT   "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT   the genetic complexity of a primitive plastid.";
RL   (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL   Schwemmler W. (eds.);
RL   Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL   (1997).
```

```
RN   [3]
RP   SEQUENCE OF 138-299 FROM N.A.
RX   MEDLINE=92201692; PubMed=1551590;
RA   Rhiel E., Stirewalt V.L., Gasparich G.E., Bryant D.A.;
RT   "The psaC genes of Synechococcus sp. PCC7002 and Cyanophora paradoxa:
RT   cloning and sequence analysis.";
RL   Gene 112:123-128(1992).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30821; AAA81303.1; -.
DR   EMBL; M86239; AAA65471.1; -.
DR   PIR; PS0372; PS0372.
DR   InterPro; IPR001626; ABC-3.
DR   Pfam; PF00950; ABC-3; 1.
KW   Hypothetical protein; Cyanelle; Transmembrane.
FT   TRANSMEM     13     33       POTENTIAL.
FT   TRANSMEM     36     56       POTENTIAL.
FT   TRANSMEM     79     99       POTENTIAL.
FT   TRANSMEM    112    132       POTENTIAL.
FT   TRANSMEM    151    171       POTENTIAL.
FT   TRANSMEM    201    221       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    267    287       POTENTIAL.
SQ   SEQUENCE   299 AA;  32817 MW;  7C8C30A20753CE66 CRC64;


  Query Match             7.2%;  Score 8;  DB 1;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 LVLALLVT 15
          ||||||||
Db    220 LVLALLVT 227


RESULT   4
Y4HA_RHISN
ID   Y4HA_RHISN     STANDARD;      PRT;   367 AA.
AC   P55471;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PUTATIVE IONIC TRANSPORTER Y4HA.
GN   Y4HA.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97305956; PubMed=9163424;
RA   Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- FUNCTION: POSSIBLE CATION TRANSPORTER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: STRONG, TO E.COLI CALCIUM/PROTON ANTIPORTER (CHAA).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000075; AAB91689.1; -.
KW   Hypothetical protein; Transport; Ion transport; Transmembrane;
KW   Plasmid.
FT   TRANSMEM     12     32       POTENTIAL.
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM     74     94       POTENTIAL.
FT   TRANSMEM    108    128       POTENTIAL.
FT   TRANSMEM    143    163       POTENTIAL.
FT   TRANSMEM    172    192       POTENTIAL.
FT   TRANSMEM    221    241       POTENTIAL.
FT   TRANSMEM    249    269       POTENTIAL.
FT   TRANSMEM    291    311       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    347    367       POTENTIAL.
SQ   SEQUENCE   367 AA;  38078 MW;  3AC0B829F2EFA2CB CRC64;


  Query Match             7.2%;  Score 8;  DB 1;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 GALLVLAL 12
          ||||||||
Db    224 GALLVLAL 231


RESULT   5
NAGA_VIBCH
ID   NAGA_VIBCH     STANDARD;      PRT;   378 AA.
AC   O32445; Q9KTA9;
DT   15-JUL-1998 (Rel. 36, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE (EC 3.5.1.25) (GLCNAC 6-P
DE   DEACETYLASE).
GN   NAGA OR VC0994.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NON-O1 / 1148A;
RX   MEDLINE=97446530; PubMed=9301118;
RA   Yamano N., Oura N., Wang J., Fujishima S.;
RT   "Cloning and sequencing of the genes for N-acetylglucosamine use that
RT   construct divergent operons (nagE-nagAC) from Vibrio cholerae
RT   non-O1.";
RL   Biosci. Biotechnol. Biochem. 61:1349-1353(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- CATALYTIC ACTIVITY: N-ACETYL-D-GLUCOSAMINE 6-PHOSPHATE + H(2)O
CC       = D-GLUCOSAMINE 6-PHOSPHATE + ACETATE.
CC   -!- PATHWAY: N-ACETYL GLUCOSAMINE UTILIZATION PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE NAGA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D87820; BAA22834.1; -.
DR   EMBL; AE004181; AAF94155.1; -.
DR   TIGR; VC0994; -.
DR   InterPro; IPR003764; NagA.
DR   Pfam; PF02612; NagA; 1.
KW   Hydrolase; Carbohydrate metabolism; Complete proteome.
FT   CONFLICT     79     79       T -> I (IN REF. 1).
SQ   SEQUENCE   378 AA;  40956 MW;  34906344A3F92A0F CRC64;


  Query Match             7.2%;  Score 8;  DB 1;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 EAVQNTVE 102
           ||||||||
Db     312 EAVQNTVE 319


RESULT   6
ITA7_RAT
ID   ITA7_RAT       STANDARD;      PRT;  1106 AA.
AC   Q63258; Q63026; Q63027;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   INTEGRIN ALPHA-7 (H36-ALPHA7).
GN   ITGA7.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA-7X1B).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=92242309; PubMed=1315319;
RA   Song W.K., Wang W., Foster R.F., Bielser D.A., Kaufman S.J.;
RT   "H36-alpha 7 is a novel integrin alpha chain that is developmentally
RT   regulated during skeletal myogenesis.";
RL   J. Cell Biol. 117:643-657(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS ALPHA-7X1A AND ALPHA-7X1C).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=94171924; PubMed=8126096;
RA   Song W.K., Wang W., Sato H., Bielser D.A., Kaufman S.J.;
RT   "Expression of alpha 7 integrin cytoplasmic domains during skeletal
RT   muscle development: alternate forms, conformational change, and
RT   homologies with serine/threonine kinases and tyrosine phosphatases.";
RL   J. Cell Sci. 106:1139-1152(1993).
RN   [3]
RP   TISSUE SPECIFITY.
RX   MEDLINE=96197133; PubMed=8626012;
RA   Martin P.T., Kaufman S.J., Kramer R.H., Sanes J.R.;
RT   "Synaptic integrins in developing, adult, and mutant muscle: selective
RT   association of alpha1, alpha7A, and alpha7B integrins with the
RT   neuromuscular junction.";
RL   Dev. Biol. 174:125-139(1996).
CC   -!- FUNCTION: INTEGRIN ALPHA-7/BETA-1 IS THE PRIMARY LAMININ RECEPTOR
CC       ON SKELETAL MYOBLASTS AND ADULT MYOFIBERS. DURING MYOGENIC
CC       DIFFERENTIATION, IT MAY INDUCE CHANGES IN THE SHAPE AND MOBILITY
CC       OF MYOBLASTS, AND FACILITATE THEIR LOCALIZATION AT LAMININ-RICH
CC       SITES OF SECONDARY FIBER FORMATION. INVOLVED IN THE MAINTENANCE OF
CC       THE MYOFIBERS CYTOARCHITECTURE AS WELL AS FOR THEIR ANCHORAGE,
CC       VIABILITY AND FUNCTIONAL INTEGRITY.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC       SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC       DISULFIDE BOND. ALPHA-7 ASSOCIATES WITH BETA-1.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; ALPHA-7X1A, ALPHA-7X1B
CC       (SHOWN HERE) AND ALPHA-7X1C; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN SKELETAL AND CARDIAC MUSCLE.
CC       EXPRESSED IN REPLICATING MYOBLASTS. IN DIFFERENTIATED MUSCLE
CC       FIBERS LOCALIZES BETWEEN FIBERS AND THE SURROUNDING MATRIX.
CC       ISOFORMS A AND B ARE EXPRESSED AT MYOTENDINOUS AND NEUROMUSCULAR
CC       JUNCTIONS; ISOFORM C IS EXPRESSED AT NEUROMUSCULAR JUNCTIONS AND
CC       AT EXTRASYNAPTIC SITES.
CC   -!- DEVELOPMENTAL STAGE: ISOFORMS ARE DEVELOPMENTALLY REGULATED DURING
CC       THE FORMATION OF SKELETAL MUSCLE. ISOFORMS A AND C ARE INDUCED
CC       UPON TERMINAL MYOGENIC DIFFERENTIATION; ISOFORM B IS PRESENT
CC       EARLIER IN REPLICATING CELLS AND DIMINISHES UPON DIFFERENTIATION.
CC   -!- PTM: ADP-RIBOSYLATED ON AT LEAST TWO SITES IN THE EXTRACELLULAR
CC       DOMAIN IN SKELETAL MYOTUBES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65036; CAA46170.1; -.
DR   EMBL; X74293; CAA52346.1; -.
DR   EMBL; X74294; CAA52347.1; -.
DR   HSSP; P11215; 1A8X.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   Pfam; PF00357; integrin_A; 1.
DR   PRINTS; PR01185; INTEGRINA.
DR   SMART; SM00191; Int_alpha; 5.
DR   PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW   Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW   Extracellular matrix; Cytoskeleton; Repeat;
KW   Alternative splicing; ADP-ribosylation; Calcium.
FT   CHAIN         1    882       INTEGRIN ALPHA-7 HEAVY CHAIN (POTENTIAL).
FT   CHAIN       886   1106       INTEGRIN ALPHA-7 LIGHT CHAIN (POTENTIAL).
FT   DOMAIN        1   1003       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1004   1029       POTENTIAL.
FT   DOMAIN     1030   1106       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       17     55       FG-GAP 1.
FT   REPEAT       90    122       FG-GAP 2.
FT   REPEAT      162    194       FG-GAP 3.
FT   REPEAT      233    269       FG-GAP 4.
FT   REPEAT      290    328       FG-GAP 5.
FT   REPEAT      351    387       FG-GAP 6.
FT   REPEAT      409    449       FG-GAP 7.
FT   CA_BIND     300    308       POTENTIAL.
FT   CA_BIND     362    370       POTENTIAL.
FT   CA_BIND     420    428       POTENTIAL.
FT   DOMAIN      880    885       POLY-ARG.
FT   SITE       1032   1036       GFFKR MOTIF.
FT   DOMAIN     1082   1101       3 X 4 AA REPEATS OF D-X-H-P.
FT   REPEAT     1082   1085       1.
FT   REPEAT     1090   1093       2.
FT   REPEAT     1098   1101       3.
FT   DISULFID     61     71       BY SIMILARITY.
FT   DISULFID    108    131       BY SIMILARITY.
FT   DISULFID    152    165       BY SIMILARITY.
FT   DISULFID    467    474       BY SIMILARITY.
FT   DISULFID    480    543       BY SIMILARITY.
FT   DISULFID    609    615       BY SIMILARITY.
FT   DISULFID    708    719       BY SIMILARITY.
FT   DISULFID    866    920       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    926    931       BY SIMILARITY.
FT   CARBOHYD    713    713       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    915    915       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    950    950       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    970    970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1031   1106       LGFFKRAKHPEATVPQYHAVKILREDRQQFKEEKTGTIQRS
FT                                NWGNSQWEGSDAHPILAADWHPELGPDGHPVSVTA -> CG
FT                                FFRRNSPSSSFPANYHRAHLAVQPSAMEAGGPGTVGWDSSS
FT                                GRSTLRPLYPSTTQ (IN ISOFORM ALPHA-7X1A).
FT   VARSPLIC   1036   1106       RAKHPEATVPQYHAVKILREDRQQFKEEKTGTIQRSNWGNS
FT                                QWEGSDAHPILAADWHPELGPDGHPVSVTA -> CAVPAQR
FT                                ILSIY (IN ISOFORM ALPHA-7X1C).
SQ   SEQUENCE   1106 AA;  121101 MW;  21B2A187837E01F6 CRC64;


  Query Match             7.2%;  Score 8;  DB 1;  Length 1106;
  Best Local Similarity   100.0%;  Pred. No. 5.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALLV 14
           ||||||||
Db    1018 LLVLALLV 1025


RESULT   7
ITA7_MOUSE
ID   ITA7_MOUSE     STANDARD;      PRT;  1179 AA.
AC   Q61738; P70350; O88732; Q61737; O88731; Q61741;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   INTEGRIN ALPHA-7 PRECURSOR.
GN   ITGA7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS ALPHA-7A; ALPHA-7X2B AND ALPHA-7X1X2).
RC   STRAIN=BALB/C; TISSUE=Heart;
RX   MEDLINE=94075378; PubMed=8253814;
RA   Ziober B.L., Vu M.P., Waleh N., Crawford J., Lin C.-S., Kramer R.H.;
RT   "Alternative extracellular and cytoplasmic domains of the integrin
RT   alpha 7 subunit are differentially expressed during development.";
RL   J. Biol. Chem. 268:26773-26783(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS ALPHA-7X1X2A AND ALPHA-7X1X2B).
RC   STRAIN=129/SV;
RA   Saher G., Echtermeyer F., Beier D.R., Poeschl E., Mayer U.;
RT   "Genomic organization and chromosomal localization of the mouse
RT   integrin alpha7 gene.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-70 FROM N.A.
RC   STRAIN=C57BL/6 x CBA;
RX   MEDLINE=96394366; PubMed=8798472;
RA   Ziober B.L., Kramer R.H.;
RT   "Identification and characterization of the cell type-specific and
RT   developmentally regulated alpha7 integrin gene promoter.";
RL   J. Biol. Chem. 271:22915-22922(1996).
RN   [4]
RP   SEQUENCE OF 34-58.
RC   TISSUE=Melanoma;
RX   MEDLINE=92198982; PubMed=1839357;
RA   Kramer R.H., Vu M.P., Cheng Y.F., Ramos D.M., Timpl R., Waleh N.;
RT   "Laminin-binding integrin alpha 7 beta 1: functional characterization
RT   and expression in normal and malignant melanocytes.";
RL   Cell Regul. 2:805-817(1991).
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORMS ALPHA-7A AND ALPHA-7B).
RC   STRAIN=C57BL/6 x BALB/C;
RX   MEDLINE=93366824; PubMed=8360188;
RA   Collo G., Starr L., Quaranta V.;
RT   "A new isoform of the laminin receptor integrin alpha 7 beta 1 is
RT   developmentally regulated in skeletal muscle.";
RL   J. Biol. Chem. 268:19019-19024(1993).
RN   [6]
RP   TISSUE SPECIFITY.
RX   MEDLINE=96197133; PubMed=8626012;
RA   Martin P.T., Kaufman S.J., Kramer R.H., Sanes J.R.;
RT   "Synaptic integrins in developing, adult, and mutant muscle: selective
RT   association of alpha1, alpha7A, and alpha7B integrins with the
RT   neuromuscular junction.";
RL   Dev. Biol. 174:125-139(1996).
RN   [7]
RP   FUNCTION.
RX   MEDLINE=98016417; PubMed=9354797;
RA   Mayer U., Saher G., Fassler R., Bornemann A., Echtermeyer F.,
RA   von der Mark H., Miosge N., Poeschl E., von der Mark K.;
RT   "Absence of integrin alpha 7 causes a novel form of muscular
RT   dystrophy.";
RL   Nat. Genet. 17:318-323(1997).
RN   [8]
RP   ADP-RYBOSYLATION.
RX   MEDLINE=95238432; PubMed=7721841;
RA   Zolkiewska A., Moss J.;
RT   "Processing of ADP-ribosylated integrin alpha 7 in skeletal muscle
RT   myotubes.";
RL   J. Biol. Chem. 270:9227-9233(1995).
CC   -!- FUNCTION: INTEGRIN ALPHA-7/BETA-1 IS THE PRIMARY LAMININ RECEPTOR
CC       ON SKELETAL MYOBLASTS AND ADULT MYOFIBERS. DURING MYOGENIC
CC       DIFFERENTIATION, IT MAY INDUCE CHANGES IN THE SHAPE AND MOBILITY
CC       OF MYOBLASTS, AND FACILITATE THEIR LOCALIZATION AT LAMININ-RICH
CC       SITES OF SECONDARY FIBER FORMATION. INVOLVED IN THE MAINTENANCE OF
CC       THE MYOFIBERS CYTOARCHITECTURE AS WELL AS FOR THEIR ANCHORAGE,
CC       VIABILITY AND FUNCTIONAL INTEGRITY. MICE CARRYING A ITGA7 NULL
CC       ALLELE ARE VIABLE AND FERTILE, BUT SHOW PROGRESSIVE MUSCULAR
CC       DYSTROPHY STARTING SOON AFTER BIRTH, BUT WITH A DISTINCT
CC       VARIABILITY IN DIFFERENT MUSCLE TYPES.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC       SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC       DISULFIDE BOND. ALPHA-7 ASSOCIATES WITH BETA-1.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; ALPHA-7X1A, ALPHA-7X1B,
CC       ALPHA-7X2A, ALPHA-7X2B, ALPHA-7X1X2A AND ALPHA-7X1X2B (SHOWN
CC       HERE); MAY BE PRODUCED BY ALTERNATIVE SPLICING. THERE IS A
CC       COMBINATION OF AT LEAST FOUR ALTERNATIVE SPLICED DOMAINS, TWO
CC       EXTRACELLULAR (X1 AND X2) AND TWO CYTOPLASMIC (A AND B). A THIRD
CC       POTENTIAL ALTERNATIVE SPLICED CYTOPLASMIC DOMAIN (C) DOESN'T
CC       APPEAR TO BE EXPRESSED. SO FAR DETECTED ARE ISOFORMS ALPHA-7X1A,
CC       ALPHA-7X1B AND ALPHA-7X2B.
CC   -!- TISSUE SPECIFICITY: ISOFORMS CONTAINING SEGMENT X2 ARE FOUND IN
CC       ADULT HEART, LUNG AND SKELETAL MUSCLE. ISOFORMS CONTAINING SEGMENT
CC       X1 ARE EXPRESSED IN ADULT HEART, LUNG AND IN PROLIFERATING
CC       SKELETAL MYOBLASTS BUT NOT IN ADULT SKELETAL MUSCLE. ISOFORMS
CC       CONTAINING SEGMENT A ARE EXCLUSIVELY FOUND IN SKELETAL MUSCLE.
CC       ISOFORMS CONTAINING SEGMENT B ARE WIDELY EXPRESSED. IN MUSCLE
CC       FIBERS ISOFORMS CONTAINING SEGMENT A AND B ARE EXPRESSED AT
CC       MYOTENDINOUS AND NEUROMUSCULAR JUNCTIONS; ISOFORMS CONTAINING
CC       SEGMENT C ARE EXPRESSED AT NEUROMUSCULAR JUNCTIONS AND AT
CC       EXTRASYNAPTIC SITES.
CC   -!- DEVELOPMENTAL STAGE: ISOFORMS ARE DEVELOPMENTALLY REGULATED DURING
CC       FORMATION OF SKELETAL MUSCLE. UNDIFFERENTIATED (REPLICATING)
CC       MYOBLASTS EXPRESS ISOFORMS CONTAINING SEGMENT B ONLY, WHEREAS
CC       DIFFERENTIATED MYOBLASTS EXPRESS ISOFORMS CONTAINING SEGMENTS A OR
CC       B.
CC   -!- PTM: ADP-RIBOSYLATED ON AT LEAST TWO SITES OF THE EXTRACELLULAR
CC       DOMAIN IN SKELETAL MYOTUBES (IN VITRO).
CC   -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L23423; AAA16600.1; -.
```

```
DR   EMBL; U60419; AAC52772.1; -.
DR   EMBL; Y12380; CAA73023.1; -.
DR   EMBL; Y12383; CAA73023.1; JOINED.
DR   EMBL; Y12384; CAA73023.1; JOINED.
DR   EMBL; Y12385; CAA73023.1; JOINED.
DR   EMBL; Y12386; CAA73023.1; JOINED.
DR   EMBL; Y12387; CAA73023.1; JOINED.
DR   EMBL; Y12388; CAA73023.1; JOINED.
DR   EMBL; Y12389; CAA73023.1; JOINED.
DR   EMBL; Y12390; CAA73023.1; JOINED.
DR   EMBL; Y12382; CAA73023.1; JOINED.
DR   EMBL; L23422; AAA16599.1; -.
DR   EMBL; Y12380; CAA73024.1; -.
DR   EMBL; Y12383; CAA73024.1; JOINED.
DR   EMBL; Y12384; CAA73024.1; JOINED.
DR   EMBL; Y12385; CAA73024.1; JOINED.
DR   EMBL; Y12386; CAA73024.1; JOINED.
DR   EMBL; Y12387; CAA73024.1; JOINED.
DR   EMBL; Y12388; CAA73024.1; JOINED.
DR   EMBL; Y12389; CAA73024.1; JOINED.
DR   EMBL; Y12390; CAA73024.1; JOINED.
DR   EMBL; Y12381; CAA73024.1; JOINED.
DR   EMBL; Y12382; CAA73024.1; JOINED.
DR   EMBL; L23421; AAA16598.1; -.
DR   EMBL; L16544; -; NOT_ANNOTATED_CDS.
DR   HSSP; P11215; 1A8X.
DR   MGD; MGI:102700; Itga7.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   Pfam; PF00357; integrin_A; 2.
DR   PRINTS; PR01185; INTEGRINA.
DR   SMART; SM00191; Int_alpha; 5.
DR   PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW   Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW   Signal; Extracellular matrix; Cytoskeleton; Repeat;
KW   Alternative splicing; ADP-ribosylation; Calcium.
FT   SIGNAL       1     33
FT   CHAIN       34   1179       INTEGRIN ALPHA-7.
FT   CHAIN       34    955       INTEGRIN ALPHA-7 HEAVY CHAIN (POTENTIAL).
FT   CHAIN      959   1179       INTEGRIN ALPHA-7 LIGHT CHAIN (POTENTIAL).
FT   DOMAIN      34   1076       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1077   1102       POTENTIAL.
FT   DOMAIN    1103   1179       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      50     87       FG-GAP 1.
FT   REPEAT     122    154       FG-GAP 2.
FT   REPEAT     194    226       FG-GAP 3.
FT   REPEAT     305    341       FG-GAP 4.
FT   REPEAT     362    400       FG-GAP 5.
FT   REPEAT     423    459       FG-GAP 6.
FT   REPEAT     481    521       FG-GAP 7.
FT   CA_BIND    372    380       POTENTIAL.
FT   CA_BIND    434    442       POTENTIAL.
FT   CA_BIND    492    500       POTENTIAL.
FT   DOMAIN     953    958       POLY-ARG.
FT   SITE      1105   1109       GFFKR MOTIF.
FT   DOMAIN    1155   1174       3 X 4 AA REPEATS OF D-X-H-P.
FT   REPEAT    1155   1158       1.
FT   REPEAT    1163   1166       2.
FT   REPEAT    1171   1174       3.
FT   DISULFID    94    103       BY SIMILARITY.
FT   DISULFID   140    163       BY SIMILARITY.
FT   DISULFID   184    197       BY SIMILARITY.
FT   DISULFID   539    546       BY SIMILARITY.
FT   DISULFID   552    615       BY SIMILARITY.
FT   DISULFID   681    687       BY SIMILARITY.
FT   DISULFID   781    792       BY SIMILARITY.
FT   DISULFID   939    993       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   999   1004       BY SIMILARITY.
FT   CARBOHYD    86     86       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   784    784       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   988    988       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1023   1023       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1043   1043       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   224    267       MISSING (IN ISOFORM ALPHA-7X2A AND
FT                               ISOFORM ALPHA-7X2B).
FT   VARSPLIC   268    307       MISSING (IN ISOFORM ALPHA-7X1A AND

  Query Match              7.2%;  Score 8;  DB 1;  Length 1179;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALLV 14
           ||||||||
Db    1091 LLVLALLV 1098


RESULT   8
ITA7_HUMAN
ID   ITA7_HUMAN     STANDARD;      PRT;  1181 AA.
AC   Q13683; O43197; Q9UEV2; Q9UET0; Q9NY89;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   INTEGRIN ALPHA-7 PRECURSOR.
GN   ITGA7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS ALPHA-7X2B AND ALPHA-7X2DB).
RC   TISSUE=Fetal heart, and Osteoblast;
RX   MEDLINE=98139911; PubMed=9473524;
RA   Leung E., Lim S.P., Berg R., Yang Y., Ni J., Wang S.-X.,
RA   Krissansen G.W.;
RT   "A novel extracellular domain variant of the human integrin alpha 7
RT   subunit generated by alternative intron splicing.";
RL   Biochem. Biophys. Res. Commun. 243:317-325(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA-7X2B).
RX   MEDLINE=98250181; PubMed=9590299;
RA   Hayashi Y.K., Chou F.-L., Engvall E., Ogawa M., Matsuda C.,
RA   Hirabayashi S., Yokochi K., Ziober B.L., Kramer R.H., Kaufman S.J.,
RA   Ozawa E., Goto Y.-I., Nonaka I., Tsukahara T., Wang J.Z.,
RA   Hoffman E.P., Arahata K.;
RT   "Mutations in the integrin alpha7 gene cause congenital myopathy.";
RL   Nat. Genet. 19:94-97(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA-7X2B).
RA   Vizirianakis I.S., Ziober B.L., Kramer R.H.;
RT   "Cloning of human integrin alpha-7 cDNA.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=99333684; PubMed=10403775;
RA   Vignier N., Moghadaszadeh B., Gary F., Beckmann J., Mayer U.,
RA   Guicheney P.;
RT   "Structure, genetic localization, and identification of the cardiac
RT   and skeletal muscle transcripts of the human integrin alpha7 gene
RT   (ITGA7).";
RL   Biochem. Biophys. Res. Commun. 260:357-364(1999).
RN   [5]
RP   SEQUENCE OF 34-45.
RC   TISSUE=Melanoma;
RX   MEDLINE=92198982; PubMed=1839357;
RA   Kramer R.H., Vu M.P., Cheng Y.F., Ramos D.M., Timpl R., Waleh N.;
RT   "Laminin-binding integrin alpha 7 beta 1: functional characterization
RT   and expression in normal and malignant melanocytes.";
RL   Cell Regul. 2:805-817(1991).
RN   [6]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM ALPHA-7X2).
RC   TISSUE=Heart;
RX   MEDLINE=94075378; PubMed=8253814;
RA   Ziober B.L., Vu M.P., Waleh N., Crawford J., Lin C.-S., Kramer R.H.;
RT   "Alternative extracellular and cytoplasmic domains of the integrin
```

```
RT   alpha 7 subunit are differentially expressed during development.";
RL   J. Biol. Chem. 268:26773-26783(1993).
RN   [7]
RP   SEQUENCE OF 1105-1181 FROM N.A. (ISOFORM ALPHA-7B).
RC   TISSUE=Fetal muscle;
RX   MEDLINE=94171924; PubMed=8126096;
RA   Song W.K., Wang W., Sato H., Bielser D.A., Kaufman S.J.;
RT   "Expression of alpha 7 integrin cytoplasmic domains during skeletal
RT   muscle development: alternate forms, conformational change, and
RT   homologies with serine/threonine kinases and tyrosine phosphatases.";
RL   J. Cell Sci. 106:1139-1152(1993).
RN   [8]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM ALPHA-7A).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=98012902; PubMed=9352853;
RA   Basora N., Vachon P.H., Herring-Gillam F.E., Perreault N.,
RA   Beaulieu J.-F.;
RT   "Relation between integrin alpha7Bbetal expression in human intestinal
RT   cells and enterocytic differentiation.";
RL   Gastroenterology 113:1510-1521(1997).
RN   [9]
RP   FUNCTION.
RX   MEDLINE=97453229; PubMed=9307969;
RA   Ziober B.L., Chen Y.Q., Kramer R.H.;
RT   "The laminin-binding activity of the alpha 7 integrin receptor is
RT   defined by developmentally regulated splicing in the extracellular
RT   domain.";
RL   Mol. Biol. Cell 8:1723-1734(1997).
RN   [10]
RP   FUNCTION.
RX   MEDLINE=20160722; PubMed=10694445;
RA   Schoeber S., Mielenz D., Echtermeyer F., Hapke S., Poeschl E.,
RA   von der Mark H., Moch H., von der Mark K.;
RT   "The role of extracellular and cytoplasmic splice domains of
RT   alpha7-integrin in cell adhesion and migration on laminins.";
RL   Exp. Cell Res. 255:303-313(2000).
RN   [11]
RP   TISSUE SPECIFITY.
RX   MEDLINE=96197133; PubMed=8626012;
RA   Martin P.T., Kaufman S.J., Kramer R.H., Sanes J.R.;
RT   "Synaptic integrins in developing, adult, and mutant muscle: selective
RT   association of alpha1, alpha7A, and alpha7B integrins with the
RT   neuromuscular junction.";
RL   Dev. Biol. 174:125-139(1996).
CC   -!- FUNCTION: INTEGRIN ALPHA-7/BETA-1 IS THE PRIMARY LAMININ RECEPTOR
CC       ON SKELETAL MYOBLASTS AND ADULT MYOFIBERS. DURING MYOGENIC
CC       DIFFERENTIATION, IT MAY INDUCE CHANGES IN THE SHAPE AND MOBILITY
CC       OF MYOBLASTS, AND FACILITATE THEIR LOCALIZATION AT LAMININ-RICH
CC       SITES OF SECONDARY FIBER FORMATION. IT IS INVOLVED IN THE
CC       MAINTENANCE OF THE MYOFIBERS CYTOARCHITECTURE AS WELL AS FOR THEIR
CC       ANCHORAGE, VIABILITY AND FUNCTIONAL INTEGRITY. ISOFORMS ALPHA-7X2B
CC       AND ALPHA-7X1B PROMOTE MYOBLAST MIGRATION ON LAMININ 1 AND LAMININ
CC       2/4, BUT ALPHA-7X1B IS LESS ACTIVE ON LAMININ 1 (IN VITRO).
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC       SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC       DISULFIDE BOND. ALPHA-7 ASSOCIATES WITH BETA-1.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 12 ISOFORMS; ALPHA-7X1A, ALPHA-
CC       7X1B, ALPHA-7X1DA, ALPHA-7X1DB, ALPHA-7X2A, ALPHA-7X2B, ALPHA-
CC       7X2DA, ALPHA-7X2DB, ALPHA-7X1X2A, ALPHA-7X1X2B (SHOWN HERE),
CC       ALPHA-7X1X2DA AND ALPHA-7X1X2DB; MAY BE PRODUCED BY ALTERNATIVE
CC       SPLICING. THERE IS A COMBINATION OF AT LEAST FIVE ALTERNATIVE
CC       SPLICED DOMAINS, THREE EXTRACELLULAR (X1, X2 AND D) AND TWO
CC       CYTOPLASMIC (A AND B). A THIRD POTENTIAL ALTERNATIVE SPLICED
CC       CYTOPLASMIC DOMAIN (C) DOES NOT APPEAR TO BE EXPRESSED. IN VITRO
CC       GENERATED ISOFORM X2C SHOWS FUNCTION. SO FAR DETECTED ARE ISOFORMS
CC       ALPHA-7X1A, ALPHA-7X2B (MAJOR), ALPHA-7X2DB (MINOR) AND MINOR
CC       ISOFORMS CONTAINING SEGMENT X1X2.
CC   -!- TISSUE SPECIFICITY: ISOFORMS CONTAINING SEGMENT A ARE
CC       PREDOMINANTLY EXPRESSED IN SKELETAL MUSCLE. ISOFORMS CONTAINING
CC       SEGMENT B ARE ABUNDANTLY EXPRESSED IN SKELETAL MUSCLE, MODERATELY
CC       IN CARDIAC MUSCLE, SMALL INTESTINE, COLON, OVARY AND PROSTATE AND
CC       WEAKLY IN LUNG AND TESTES. ISOFORMS CONTAINING SEGMENT X2D ARE
CC       EXPRESSED AT LOW LEVELS IN FETAL AND ADULT SKELETAL MUSCLE AND IN
CC       CARDIAC MUSCLE, BUT ARE NOT DETECTED IN MYOBLASTS AND MYOTUBES. IN
CC       MUSCLE FIBERS ISOFORMS CONTAINING SEGMENT A AND B ARE EXPRESSED AT
CC       MYOTENDINOUS AND NEUROMUSCULAR JUNCTIONS; ISOFORMS CONTAINING
CC       SEGMENT C ARE EXPRESSED AT NEUROMUSCULAR JUNCTIONS AND AT
CC       EXTRASYNAPTIC SITES. ISOFORMS CONTAINING SEGMENTS X1 OR X2 OR, AT
CC       LOW LEVELS, X1X2 ARE EXPRESSED IN FETAL AND ADULT SKELETAL MUSCLE
CC       (MYOBLASTS AND MYOTUBES) AND CARDIAC MUSCLE.
CC   -!- DEVELOPMENTAL STAGE: IN RENEWING INTESTINAL EPITHELIUM, EXPRESSION
CC       OF ISOFORMS CONTAINING SEGMENT B CORRELATES WITH THE ONSET OF
CC       ENTEROCYTIC DIFFERENTIATION.
CC   -!- PTM: ADP-RIBOSYLATED ON AT LEAST TWO SITES OF THE EXTRACELLULAR
CC       DOMAIN IN SKELETAL MYOTUBES (BY SIMILARITY).
CC   -!- DISEASE: DEFECTS IN ITGA7 ARE ASSOCIATED WITH A FORM OF CONGENITAL
CC       MYOPATHY; A GROUP OF HETEROGENEOUS MUSCLE DISORDERS WHICH ARE
CC       THOUGHT TO RESULT FROM ABNORMAL MUSCLE DEVELOPMENT. MUSCLE
CC       WEAKNESS IS EITHER NON-PROGRESSIVE OR SLOWLY PROGRESSIVE AND
CC       APPARENT FROM BIRTH OR EARLY INFANCY.
CC   -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF032108; AAC39708.1; -.
DR   EMBL; AF052050; AAC18968.1; -.
DR   EMBL; AF072132; AAC80458.1; -.
DR   EMBL; AJ228836; CAB41534.1; -.
DR   EMBL; AJ228837; CAB41534.1; JOINED.
DR   EMBL; AJ228838; CAB41534.1; JOINED.
DR   EMBL; AJ228839; CAB41534.1; JOINED.
DR   EMBL; AJ228840; CAB41534.1; JOINED.
DR   EMBL; AJ228842; CAB41534.1; JOINED.
DR   EMBL; AJ228843; CAB41534.1; JOINED.
DR   EMBL; AJ228844; CAB41534.1; JOINED.
DR   EMBL; AJ228845; CAB41534.1; JOINED.
DR   EMBL; AJ228846; CAB41534.1; JOINED.
DR   EMBL; AJ228847; CAB41534.1; JOINED.
DR   EMBL; AJ228848; CAB41534.1; JOINED.
DR   EMBL; AJ228849; CAB41534.1; JOINED.
DR   EMBL; AJ228850; CAB41534.1; JOINED.
DR   EMBL; AJ228851; CAB41534.1; JOINED.
DR   EMBL; AJ228852; CAB41534.1; JOINED.
DR   EMBL; AJ228853; CAB41534.1; JOINED.
DR   EMBL; AJ228854; CAB41534.1; JOINED.
DR   EMBL; AJ228855; CAB41534.1; JOINED.
DR   EMBL; AJ228856; CAB41534.1; JOINED.
DR   EMBL; AJ228857; CAB41534.1; JOINED.
DR   EMBL; AJ228858; CAB41534.1; JOINED.
DR   EMBL; AJ228859; CAB41534.1; JOINED.
DR   EMBL; AJ228860; CAB41534.1; JOINED.
DR   EMBL; AJ228862; CAB41534.1; JOINED.
DR   EMBL; AJ228836; CAB41535.1; -.
DR   EMBL; AJ228837; CAB41535.1; JOINED.
DR   EMBL; AJ228838; CAB41535.1; JOINED.
DR   EMBL; AJ228839; CAB41535.1; JOINED.
DR   EMBL; AJ228841; CAB41535.1; JOINED.
DR   EMBL; AJ228842; CAB41535.1; JOINED.
DR   EMBL; AJ228843; CAB41535.1; JOINED.
DR   EMBL; AJ228844; CAB41535.1; JOINED.
DR   EMBL; AJ228845; CAB41535.1; JOINED.
DR   EMBL; AJ228846; CAB41535.1; JOINED.
DR   EMBL; AJ228847; CAB41535.1; JOINED.
DR   EMBL; AJ228848; CAB41535.1; JOINED.

  Query Match             7.2%;  Score 8;  DB 1;  Length 1181;
  Best Local Similarity   100.0%;  Pred. No. 6.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       7 LLVLALLV 14
           ||||||||
Db    1093 LLVLALLV 1100


RESULT   9
YEDD_SALTY
ID   YEDD_SALTY     STANDARD;      PRT;   141 AA.
AC   Q06399;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN YEDD.
GN   YEDD.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SJW1103;
RX   MEDLINE=93381452; PubMed=8371104;
RA   Raha M., Kihara M., Kawagishi I., Macnab R.M.;
RT   "Organization of the Escherichia coli and Salmonella typhimurium
RT   chromosomes between flagellar regions IIIa and IIIb, including a
RT   large non-coding region.";
RL   J. Gen. Microbiol. 139:1401-1407(1993).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13280; AAA71971.1; -.
DR   StyGene; SG10420; yedD.
KW   Hypothetical protein.
SQ   SEQUENCE   141 AA;  15465 MW;  A37E9E11ADCB8DD9 CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 8.9;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 GALLVLA 11
           |||||||
Db       8 GALLVLA 14


RESULT  10
LYC2_PIG
ID   LYC2_PIG       STANDARD;      PRT;   146 AA.
AC   P12068;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   LYSOZYME C-2 PRECURSOR (EC 3.2.1.17) (1,4-BETA-N-ACETYLMURAMIDASE C).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96292525; PubMed=8728388;
RA   Yu M., Irwin D.M.;
RT   "Evolution of stomach lysozyme: the pig lysozyme gene.";
RL   Mol. Phylogenet. Evol. 5:298-308(1996).
RN   [2]
RP   SEQUENCE OF 19-146.
RC   TISSUE=Stomach;
RX   MEDLINE=89362490; PubMed=2504928;
RA   Jolles J., Jolles P., Bowman B.H., Prager E.M., Stewart C.-B.,
RA   Wilson A.C.;
RT   "Episodic evolution in the stomach lysozymes of ruminants.";
RL   J. Mol. Evol. 28:528-535(1989).
CC   -!- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION; THOSE
CC       IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-
CC       MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOAGENTS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN
CC       N-ACETYL-D-GLUCOSAMINE AND N-ACETYLMURAMIC ACID IN PEPTIDOGLYCAN
CC       HETEROPOLYMERS OF THE PROKARYOTES CELL WALLS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: LYSOZYMES C ARE CAPABLE OF BOTH HYDROLYSIS &
CC       TRANSGLYCOSYLATION; THEY SHOW ALSO A SLIGHT ESTERASE ACTIVITY.
CC       THEY ACT RAPIDLY ON BOTH PEPTIDE-SUBSTITUTED & UNSUBSTITUTED
CC       PEPTIDOGLYCAN &, SLOWLY, ON CHITIN OLIGOSACCHARIDES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28757; AAB16862.1; -.
DR   HSSP; P00695; 2LHM.
DR   InterPro; IPR001916; Lactalbmn_lysozyme.
DR   Pfam; PF00062; lys; 1.
DR   PRINTS; PR00135; LYZLACT.
DR   PRINTS; PR00137; LYSOZYME.
DR   SMART; SM00263; LYZ1; 1.
DR   PROSITE; PS00128; LACTALBUMIN_LYSOZYME; 1.
KW   Hydrolase; Glycosidase; Bacteriolytic enzyme; Digestion; Stomach;
KW   Signal.
FT   SIGNAL        1     18
FT   CHAIN        19    146       LYSOZYME C-2.
FT   DISULFID     24    144       BY SIMILARITY.
FT   DISULFID     48    132       BY SIMILARITY.
FT   DISULFID     81     97       BY SIMILARITY.
FT   DISULFID     93    111       BY SIMILARITY.
FT   ACT_SITE     53     53       BY SIMILARITY.
FT   ACT_SITE     69     69       BY SIMILARITY.
SQ   SEQUENCE   146 AA;  16484 MW;  C2986F5CAAAF49F4 CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 9.2;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALL 13
           |||||||
Db       4 LLVLALL 10


RESULT  11
LYC3_PIG
ID   LYC3_PIG       STANDARD;      PRT;   148 AA.
AC   P12069;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   LYSOZYME C-3 PRECURSOR (EC 3.2.1.17) (1,4-BETA-N-ACETYLMURAMIDASE C).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Echetebu Z.O., Nevils M., Bixby J.A., Roberts R.M., Trout W.E.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
```

```
RP   SEQUENCE OF 19-148.
RC   TISSUE=Stomach;
RX   MEDLINE=89362490; PubMed=2504928;
RA   Jolles J., Jolles P., Bowman B.H., Prager E.M., Stewart C.-B.,
RA   Wilson A.C.;
RT   "Episodic evolution in the stomach lysozymes of ruminants.";
RL   J. Mol. Evol. 28:528-535(1989).
CC   -!- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION; THOSE
CC       IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-
CC       MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOAGENTS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN
CC       N-ACETYL-D-GLUCOSAMINE AND N-ACETYLMURAMIC ACID IN PEPTIDOGLYCAN
CC       HETEROPOLYMERS OF THE PROKARYOTES CELL WALLS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: LYSOZYMES C ARE CAPABLE OF BOTH HYDROLYSIS &
CC       TRANSGLYCOSYLATION; THEY SHOW ALSO A SLIGHT ESTERASE ACTIVITY.
CC       THEY ACT RAPIDLY ON BOTH PEPTIDE-SUBSTITUTED & UNSUBSTITUTED
CC       PEPTIDOGLYCAN &, SLOWLY, ON CHITIN OLIGOSACCHARIDES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U44435; AAA86644.1; -.
DR   PIR; S10047; S10047.
DR   HSSP; P00695; 2LHM.
DR   InterPro; IPR001916; Lactalbmn_lysozyme.
DR   Pfam; PF00062; lys; 1.
DR   PRINTS; PR00135; LYZLACT.
DR   PRINTS; PR00137; LYSOZYME.
DR   SMART; SM00263; LYZ1; 1.
DR   PROSITE; PS00128; LACTALBUMIN_LYSOZYME; 1.
KW   Hydrolase; Glycosidase; Bacteriolytic enzyme; Digestion; Stomach;
KW   Signal.
FT   SIGNAL        1     18
FT   CHAIN        19    148       LYSOZYME C-3.
FT   DISULFID     24    146       BY SIMILARITY.
FT   DISULFID     48    134       BY SIMILARITY.
FT   DISULFID     83     99       BY SIMILARITY.
FT   DISULFID     95    113       BY SIMILARITY.
FT   ACT_SITE     53     53       BY SIMILARITY.
FT   ACT_SITE     71     71       BY SIMILARITY.
SQ   SEQUENCE   148 AA;  16711 MW;  76AA67BFABBD64E6 CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 148;
  Best Local Similarity   100.0%;  Pred. No. 9.3;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALL 13
            |||||||
Db       4 LLVLALL 10


RESULT  12
BACR_HALS4
ID   BACR_HALS4     STANDARD;      PRT;   250 AA.
AC   O93740;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   BACTERIORHODOPSIN (BR).
GN   BOP.
OS   Halobacterium sp. (strain arg-4).
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
OX   NCBI_TaxID=160432;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99096913; PubMed=9878396;
RA   Ihara K., Umemura T., Katagiri I., Kitajima-Ihara T., Sugiyama Y.,
RA   Kimura Y., Mukohata Y.;
RT   "Evolution of the archaeal rhodopsins: evolution rate changes by gene
RT   duplication and functional differentiation.";
RL   J. Mol. Biol. 285:163-174(1999).
CC   -!- FUNCTION: BACTERIORHODOPSIN IS A LIGHT-DRIVEN PROTON PUMP.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL OPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB009620; BAA75200.1; -.
DR   HSSP; P02945; 1F50.
DR   InterPro; IPR001425; Bac_rhodopsin.
DR   Pfam; PF01036; Bac_rhodopsin; 1.
DR   PRINTS; PR00251; BACTRLOPSIN.
DR   PROSITE; PS00950; BACTERIAL_OPSIN_1; 1.
DR   PROSITE; PS00327; BACTERIAL_OPSIN_RET; 1.
KW   Ion transport; Photoreceptor; Transmembrane; Retinal protein;
KW   Hydrogen ion transport.
FT   DOMAIN        1     16       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     17     37       HELIX A (POTENTIAL).
FT   TRANSMEM     52     72       HELIX B (POTENTIAL).
FT   TRANSMEM     91    111       HELIX C (POTENTIAL).
FT   TRANSMEM    115    135       HELIX D (POTENTIAL).
FT   TRANSMEM    143    163       HELIX E (POTENTIAL).
FT   TRANSMEM    185    205       HELIX F (POTENTIAL).
FT   TRANSMEM    213    233       HELIX G (POTENTIAL).
FT   DOMAIN      234    250       CYTOPLASMIC (POTENTIAL).
FT   BINDING     225    225       RETINAL CHROMOPHORE (BY SIMILARITY).
SQ   SEQUENCE   250 AA;  27041 MW;  B72D3373506FD275 CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 LLLDLSL 48
           |||||||
Db     101 LLLDLSL 107


RESULT  13
QOX2_BACSU
ID   QOX2_BACSU     STANDARD;      PRT;   322 AA.
AC   P34957; O32281;
DT   01-FEB-1994 (Rel. 28, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   QUINOL OXIDASE POLYPEPTIDE II PRECURSOR (EC 1.9.3.-) (QUINOL OXIDASE
DE   AA3-600, SUBUNIT QOXA) (OXIDASE AA(3) SUBUNIT 2).
GN   QOXA OR IPA-37D.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=92268053; PubMed=1316894;
RA   Santana M., Kunst F., Hullo M.F., Rapoport G., Danchin A.,
RA   Glaser P.;
RT   "Molecular cloning, sequencing, and physiological characterization of
RT   the qox operon from Bacillus subtilis encoding the aa3-600 quinol
RT   oxidase.";
```

```
RL   J. Biol. Chem. 267:10225-10231(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95020537; PubMed=7934828;
RA   Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA   Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA   Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA   Rapoport G., Danchin A.;
RT   "Bacillus subtilis genome project: cloning and sequencing of the 97
RT   kb region from 325 degrees to 333 degrees.";
RL   Mol. Microbiol. 10:371-384(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 27-43, AND CHARACTERIZATION.
RC   STRAIN=W23;
RX   MEDLINE=96009139; PubMed=7575098;
RA   Lemma E., Simon J., Schagger H., Kroger A.;
RT   "Properties of the menaquinol oxidase (Qox) and of qox deletion
RT   mutants of Bacillus subtilis.";
RL   Arch. Microbiol. 163:432-438(1995).
CC   -!- FUNCTION: CATALYZES QUINOL OXIDATION WITH THE CONCOMITANT
CC       REDUCTION OF OXYGEN TO WATER. MAJOR COMPONENT FOR ENERGY
CC       CONVERSION DURING VEGETATIVE GROWTH. SUBUNIT II TRANSFERS THE
CC       ELECTRONS FROM A QUINOL TO THE BINUCLEAR CENTER OF THE CATALYTIC
CC       SUBUNIT I.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: SOME, TO MITOCHONDRIAL OR BACTERIAL COX2 SUBUNITS.
CC       BUT LACK HEME-BINDING DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86548; AAA22686.1; ALT_INIT.
DR   EMBL; X73124; CAA51593.1; ALT_INIT.
DR   EMBL; Z99123; CAB15843.1; ALT_INIT.
DR   PIR; A38129; A38129.
DR   HSSP; P18400; 1CYW.
DR   SubtiList; BG10583; qoxA.
DR   InterPro; IPR001505; COX2.
DR   InterPro; IPR002429; Cyt_c_ox_2.
DR   Pfam; PF00116; COX2; 1.
DR   PRINTS; PR01166; CYCOXIDASEII.
DR   ProDom; PD000131; COX2; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Oxidoreductase; Electron transport; Transmembrane; Signal;
KW   Lipoprotein; Complete proteome.
FT   SIGNAL       1     26
FT   CHAIN       27    322       QUINOL OXIDASE POLYPEPTIDE II.
FT   LIPID       27     27       N-ACYL DIGLYCERIDE (POTENTIAL).
FT   TRANSMEM    50     70       POTENTIAL.
FT   TRANSMEM    92    112       POTENTIAL.
FT   CONFLICT    39     39       V -> VD (IN REF. X).
SQ   SEQUENCE   322 AA;  36353 MW;  C701B14076B74A8D CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 322;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALL 13
            |||||||
Db      12 LLVLALL 18
```

```
RESULT  14
C1B2_CAVPO
ID   C1B2_CAVPO     STANDARD;      PRT;   332 AA.
AC   Q9QZZ1;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   T-CELL SURFACE GLYCOPROTEIN CD1B2 PRECURSOR (CD1-B2 ANTIGEN).
GN   CD1B2.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HARTLEY, AND NIH 2; TISSUE=Thymus;
RX   MEDLINE=20021845; PubMed=10553074;
RA   Dascher C.C., Hiromatsu K., Naylor J.W., Brauer P.P., Brown K.A.,
RA   Storey J.R., Behar S.M., Kawasaki E.S., Porcelli S.A., Brenner M.B.,
RA   LeClair K.P.;
RT   "Conservation of a CD1 multigene family in the guinea pig.";
RL   J. Immunol. 163:5478-5488(1999).
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF145484; AAF12739.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003597; Ig_c1.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00407; IGc1; 1.
KW   Glycoprotein; Signal; Transmembrane; Immunoglobulin domain;
KW   Multigene family.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    332       T-CELL SURFACE GLYCOPROTEIN CD1B2.
FT   DOMAIN      18    301       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   302    322       POTENTIAL.
FT   DOMAIN     323    332       CYTOPLASMIC (POTENTIAL).
FT   DISULFID   120    184       BY SIMILARITY.
FT   DISULFID   224    279       BY SIMILARITY.
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   146    146       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   332 AA;  37384 MW;  11D93606657ADC3A CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALL 13
            |||||||
Db       3 LLVLALL 9


RESULT  15
YE71_HAEIN
ID   YE71_HAEIN     STANDARD;      PRT;   337 AA.
AC   Q57130; O05065;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
```

```
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL ABC TRANSPORTER PERMEASE PROTEIN HI1471.
GN   HI1471.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE
CC       SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE FECCD
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32825; AAC23119.1; -.
DR   TIGR; HI1471; -.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR000522; FecCD.
DR   Pfam; PF01032; FecCD_family; 1.
DR   ProDom; PD001557; FecCD; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     66     86       POTENTIAL.
FT   TRANSMEM    109    129       POTENTIAL.
FT   TRANSMEM    147    167       POTENTIAL.
FT   TRANSMEM    196    216       POTENTIAL.
FT   TRANSMEM    247    267       POTENTIAL.
FT   TRANSMEM    281    301       POTENTIAL.
FT   TRANSMEM    307    327       POTENTIAL.
SQ   SEQUENCE   337 AA;  36530 MW;  83298CCBCF2DC1F5 CRC64;


  Query Match             6.3%;  Score 7;  DB 1;  Length 337;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      17 ALGVKMA 23
           |||||||
Db     230 ALGVKMA 236


Search completed: May  7, 2002, 12:21:10
Job time: 191 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 7, 2002, 12:16:24 ; Search time 14.31 Seconds
(without alignments)
590.872 Million cell updates/sec

Title: US-09-662-784-6
Perfect score: 111
Sequence: 1 DTMRGALLVLALLVTQALGV.....CMGEAVQNTVEDLKLNTLGR 111

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	109	98.2	109	2	C56413	major allergen Fel
2	81	73.0	107	2	JC1127	major allergen cha
3	8	7.2	108	2	S40149	integrin alpha-7C
4	8	7.2	129	2	I61187	alpha-7A integrin
5	8	7.2	139	1	C45051	lamprin 2 precurso
6	8	7.2	148	2	S40148	integrin alpha-7A
7	8	7.2	299	2	T06960	probable membrane
8	8	7.2	378	2	JC5649	N-acetylglucosamin
9	8	7.2	378	2	E82254	N-acetylglucosamin
10	8	7.2	401	2	E82951	probable MFS trans
11	8	7.2	1106	2	S38783	integrin alpha cha
12	8	7.2	1135	2	I61186	alpha-7 integrin -
13	8	7.2	1137	2	JC5950	integrin alpha-7 c
14	7	6.3	20	2	B53283	major cat allergen
15	7	6.3	73	2	T20911	hypothetical prote
16	7	6.3	77	2	D83680	hypothetical prote
17	7	6.3	87	2	G46449	hypothetical prote
18	7	6.3	103	2	G75257	hypothetical prote
19	7	6.3	198	2	G59094	hypothetical prote
20	7	6.3	203	2	B75497	hypothetical prote
21	7	6.3	244	2	D75505	hypothetical prote
22	7	6.3	250	2	T50677	bacteriorhodopsin
23	7	6.3	253	2	T46454	hypothetical prote
24	7	6.3	266	2	T36287	probable integral
25	7	6.3	268	2	A84746	probable NAM (no a
26	7	6.3	279	2	S75087	hypothetical prote
27	7	6.3	303	2	T00479	probable phosphati
28	7	6.3	321	1	E69687	bo-type ubiquinol
29	7	6.3	322	2	B83579	hypothetical prote
30	7	6.3	337	1	G64125	hemin transport pr
31	7	6.3	341	2	C72403	hypothetical prote
32	7	6.3	350	2	S75065	sensory transducti
33	7	6.3	355	2	JC1249	peroxidase (EC 1.1
34	7	6.3	364	2	S34355	peroxidase (EC 1.1
35	7	6.3	374	2	B82488	hypothetical prote
36	7	6.3	385	2	E83506	probable MFS trans
37	7	6.3	390	2	H70904	probable lprK prot
38	7	6.3	397	2	S09813	hypothetical prote
39	7	6.3	400	2	B70621	probable argD prot
40	7	6.3	400	2	T35334	probable membrane
41	7	6.3	422	2	G83503	probable MFS trans
42	7	6.3	424	2	S32488	dihydrolipoamide S
43	7	6.3	452	2	E83482	probable MFS trans
44	7	6.3	455	2	B86722	biotin carboxylase
45	7	6.3	474	2	F82412	probable aspartoki

ALIGNMENTS

RESULT 1
C56413
major allergen Fel dI chain 2 precursor - cat
C;Species: Felis silvestris catus (domestic cat)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 01-Dec-2000
C;Accession: C56413; JC1145
R;Morgenstern, J.P.; Griffith, I.J.; Brauer, A.W.; Rogers, B.L.; Bond, J.F.; Chapman,
Proc. Natl. Acad. Sci. U.S.A. 88, 9690-9694, 1991
A;Title: Amino acid sequence of Fel dI, the major allergen of the domestic cat: prote
A;Reference number: A56413; MUID:92052157
A;Accession: C56413
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-109 <MOR>
A;Cross-references: GB:M77341; NID:g163822; PIDN:AAC41616.1; PID:g163823
R;Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.
Gene 113, 263-268, 1992
A;Title: Expression and genomic structure of the genes encoding FdI, the major allerg
A;Reference number: JC1126; MUID:92241678
A;Accession: JC1145
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 18-109 <GRI>
A;Experimental source: salivary gland
C;Keywords: glycoprotein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 98.2%; Score 109; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 7.7e-101;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy      63 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KIQDCYVENGLISRVLDGLVMTTISSSKDCMGEAVQNTVEDLKLNTLGR 109
```

RESULT 2
JC1127
major allergen chain 2 precursor, short form - cat
C;Species: Felis silvestris catus (domestic cat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: JC1127
R;Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.
Gene 113, 263-268, 1992
A;Title: Expression and genomic structure of the genes encoding FdI, the major allerg
A;Reference number: JC1126; MUID:92241678

```
A;Accession: JC1127
A;Molecule type: DNA
A;Residues: 1-107 <GRI>
A;Cross-references: GB:X62478; NID:g395406; PIDN:CAA44345.1; PID:g395407
A;Experimental source: skin
C;Genetics:
A;Gene: Ch2
A;Introns: 21/1; 81/3
C;Keywords: glycoprotein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-107/Product: major allergen chain 2, short form #status predicted <MAT>
F;50/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             73.0%;  Score 81;  DB 2;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-73;
  Matches   81;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 62
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRGALLVLALLVTQALGVKMAETCPIFYDVFFAVANGNELLLDLSLTKVNATEPERTAMK 60

Qy     63 KIQDCYVENGLISRVLDGLVM 83
          |||||||||||||||||||||
Db     61 KIQDCYVENGLISRVLDGLVM 81


RESULT   3
S40149
integrin alpha-7C chain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 29-Sep-1999
C;Accession: S40149
R;Song, W.K.; Wang, W.; Sato, H.; Bielser, D.; Kaufman, S.
submitted to the EMBL Data Library, July 1993
A;Description: Expression of alpha 7 integrin cytoplasmic domains during skeletal muscle
hosphotases.
A;Reference number: S40147
A;Accession: S40149
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-108 <SON>
A;Cross-references: EMBL:X74294; NID:g437912; PIDN:CAA52347.1; PID:g437913
C;Superfamily: integrin alpha-2b chain


  Query Match              7.2%;  Score 8;  DB 2;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 LLVLALLV 14
          ||||||||
Db     79 LLVLALLV 86


RESULT   4
I61187
alpha-7A integrin - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C;Accession: I61187
R;Ziober, B.L.; Vu, M.P.; Waleh, N.; Crawford, J.; Lin, C.S.; Kramer, R.H.
J. Biol. Chem. 268, 26773-26783, 1993
A;Title: Alternative extracellular and cytoplasmic domains of the integrin alpha 7 subun
A;Reference number: A49691; MUID:94075378
A;Accession: I61187
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-129 <RES>
A;Cross-references: GB:L23421; NID:g431417; PIDN:AAA16598.1; PID:g450648
C;Superfamily: integrin alpha-2b chain


  Query Match              7.2%;  Score 8;  DB 2;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 LLVLALLV 14
          ||||||||
Db     61 LLVLALLV 68


RESULT   5
C45051
lamprin 2 precursor, long splice form - sea lamprey
N;Alternate names: lamprin L-0.9-12
N;Contains: lamprin 2 precursor, short splice form (lamprin L-0.9-10)
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 30-Apr-1993 #sequence_revision 30-Apr-1993 #text_change 22-Jun-1999
C;Accession: C45051; A45051
R;Robson, P.; Wright, G.M.; Sitarz, E.; Maiti, A.; Rawat, M.; Youson, J.H.; Keeley, F
J. Biol. Chem. 268, 1440-1447, 1993
A;Title: Characterization of lamprin, an unusual matrix protein from lamprey cartilag
A;Reference number: A45051; MUID:93123269
A;Accession: C45051
A;Molecule type: mRNA
A;Residues: 1-139 <ROB>
A;Cross-references: GB:L05925; NID:g213209; PIDN:AAA49269.1; PID:g213210
A;Note: sequence extracted from NCBI backbone (NCBIP:122170)
A;Accession: A45051
A;Molecule type: mRNA
A;Residues: 1-85,105-139 <RO2>
A;Cross-references: GB:L05924; NID:g213207; PIDN:AAA49268.1; PID:g213208
A;Note: sequence extracted from NCBI backbone (NCBIP:122166)
C;Comment: These forms are encoded by a different gene than B45051; we have arbitrari
C;Superfamily: lamprin
C;Keywords: alternative splicing; cartilage; extracellular matrix
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-139/Product: lamprin 2, long splice form #status predicted <MAT1>
F;20-85,105-139/Product: lamprin 2, short splice form #status predicted <MAT2>


  Query Match              7.2%;  Score 8;  DB 1;  Length 139;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 ALLVLALL 13
          ||||||||
Db      7 ALLVLALL 14


RESULT   6
S40148
integrin alpha-7A chain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 29-Sep-1999
C;Accession: S40148
R;Song, W.K.; Wang, W.; Sato, H.; Bielser, D.; Kaufman, S.
submitted to the EMBL Data Library, July 1993
A;Description: Expression of alpha 7 integrin cytoplasmic domains during skeletal mus
hosphotases.
A;Reference number: S40147
A;Accession: S40148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-148 <SON>
A;Cross-references: EMBL:X74293; NID:g437910; PIDN:CAA52346.1; PID:g437911
C;Superfamily: integrin alpha-2b chain


  Query Match              7.2%;  Score 8;  DB 2;  Length 148;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 LLVLALLV 14
          ||||||||
```

```
Db      79 LLVLALLV 86


RESULT   7
T06960
probable membrane protein - Cyanophora paradoxa cyanelle
N;Alternate names: hypothetical protein psaC region
C;Species: cyanelle Cyanophora paradoxa
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 24-Nov-1999
C;Accession: T06960; PS0372
R;Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A;Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A;Reference number: Z15840
A;Accession: T06960
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-299 <STI>
A;Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81303.1; PID:g1016216
A;Experimental source: strain Pringsheim LB555
R;Rhiel, E.; Stirewalt, V.L.; Gasparich, G.E.; Bryant, D.A.
Gene 112, 123-128, 1992
A;Title: The psaC genes of Synechococcus sp. PCC7002 and Cyanophora paradoxa: cloning an
A;Reference number: JS0694; MUID:92201692
A;Accession: PS0372
A;Molecule type: DNA
A;Residues: 138-299 <RHI>
A;Cross-references: GB:M86239; NID:g336634; PIDN:AAA65471.1; PID:g336637
C;Genetics:
A;Genome: cyanelle
C;Superfamily: conserved hypothetical protein HI0360
C;Keywords: cyanelle; transmembrane protein


  Query Match             7.2%;  Score 8;  DB 2;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 3.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 LVLALLVT 15
          ||||||||
Db     220 LVLALLVT 227


RESULT   8
JC5649
N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) - Vibrio cholerae (strain non-
C;Species: Vibrio cholerae
A;Note: non-O1 strain 1148A
C;Date: 28-Oct-1997 #sequence_revision 28-Oct-1997 #text_change 05-Nov-1999
C;Accession: JC5649
R;Yamano, N.; Oura, N.; Wang, J.; Fujishima, S.
Biosci. Biotechnol. Biochem. 61, 1349-1353, 1997
A;Title: Cloning and sequencing of the genes for N-acetylglucosamine use that construct
A;Reference number: JC5649; MUID:97446530
A;Accession: JC5649
A;Molecule type: DNA
A;Residues: 1-378 <YAM>
A;Cross-references: DDBJ:D87820; NID:g2541899; PIDN:BAA22834.1; PID:d1023704; PID:g2541
A;Experimental source: strain 1148A
C;Comment: This enzyme deacetylates N-acetylglucosamine, one component of chitin, to glu
C;Genetics:
A;Gene: nagA
C;Keywords: hydrolase


  Query Match             7.2%;  Score 8;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 EAVQNTVE 102
           ||||||||
Db     312 EAVQNTVE 319


RESULT   9
E82254
N-acetylglucosamine-6-phosphate deacetylase VC0994 [imported] - Vibrio cholerae (stra
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: E82254
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: E82254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <HEI>
A;Cross-references: GB:AE004181; GB:AE003852; NID:g9655454; PIDN:AAF94155.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0994
A;Map position: 1


  Query Match             7.2%;  Score 8;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 EAVQNTVE 102
           ||||||||
Db     312 EAVQNTVE 319


RESULT  10
E82951
probable MFS transporter PA5548 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E82951
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: E82951
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <STO>
A;Cross-references: GB:AE004967; GB:AE004091; NID:g9951884; PIDN:AAG08933.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA5548


  Query Match             7.2%;  Score 8;  DB 2;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LLVLALLV 14
          ||||||||
Db     281 LLVLALLV 288


RESULT  11
S38783
integrin alpha chain - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Jan-1994 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C;Accession: S38783; S23600
R;Kaufman, S.J.
submitted to the EMBL Data Library, March 1992
A;Reference number: S38783
```

A;Accession: S38783
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1106 <KAU>
A;Cross-references: EMBL:X65036; NID:g56392; PIDN:CAA46170.1; PID:g56393
R;Song, W.K.; Wang, W.; Foster, R.F.; Bielser, D.A.; Kaufman, S.J.
J. Cell Biol. 117, 643-657, 1992
A;Title: H36-alpha7 is a novel integrin alpha chain that is developmentally regulated du
A;Reference number: S23600; MUID:92242309
A;Accession: S23600
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-116,'R',118-349,'D',351-804,'R',806,'V',808-1106 <SON>
A;Cross-references: EMBL:X65036
C;Superfamily: integrin alpha-2b chain

Query Match 7.2%; Score 8; DB 2; Length 1106;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LLVLALLV 14
||||||||
Db 1018 LLVLALLV 1025

RESULT 12
I61186
alpha-7 integrin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C;Accession: I61186
R;Ziober, B.L.; Vu, M.P.; Waleh, N.; Crawford, J.; Lin, C.S.; Kramer, R.H.
J. Biol. Chem. 268, 26773-26783, 1993
A;Title: Alternative extracellular and cytoplasmic domains of the integrin alpha 7 subun
A;Reference number: A49691; MUID:94075378
A;Accession: I61186
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1135 <RES>
A;Cross-references: GB:L23423; NID:g431415; PIDN:AAA16600.1; PID:g431416
C;Superfamily: integrin alpha-2b chain

Query Match 7.2%; Score 8; DB 2; Length 1135;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LLVLALLV 14
||||||||
Db 1047 LLVLALLV 1054

RESULT 13
JC5950
integrin alpha-7 chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 31-Dec-2000
C;Accession: JC5950; JC5951; A56839; S40147
R;Leung, E.; Lim, S.P.; Berg, R.; Yang, Y.; Ni, J.; Wang, S.; Krissansen, G.W.
Biochem. Biophys. Res. Commun. 243, 317-325, 1998
A;Title: A novel extracellular domain variant of the human integrin alpha 7 subunit gene
A;Reference number: JC5950; MUID:98139911
A;Accession: JC5950
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1137 <LEU1>
A;Cross-references: GB:AF032108; NID:g2897115; PIDN:AAC39708.1; PID:g2897116
A;Accession: JC5951
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-657,733-1137 <LEU2>
A;Cross-references: GB:AF032108
R;Wang, W.; Wu, W.; Desai, T.; Ward, D.C.; Kaufman, S.J.
Genomics 26, 563-570, 1995
A;Title: Localization of the alpha7 integrin gene (ITGA7) on human chromosome 12q13:
A;Reference number: A56839
A;Accession: A56839
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 664-666,668-688,690-825,'X',827-839,'D',841-946,'H',948,'T',951,'P',953-1
A;Cross-references: EMBL:X74295; NID:g437781
R;Song, W.K.; Wang, W.; Sato, H.; Bielser, D.; Kaufman, S.
submitted to the EMBL Data Library, July 1993
A;Description: Expression of alpha 7 integrin cytoplasmic domains during skeletal mus
hosphotases.
A;Reference number: S40147
A;Accession: S40147
A;Molecule type: mRNA
A;Residues: 1061-1137 <SON>
A;Cross-references: EMBL:X74295; NID:g437781; PIDN:CAA52348.1; PID:g437782
C;Genetics:
A;Gene: GDB:ITGA7
A;Cross-references: GDB:131465; OMIM:600536
A;Map position: 12q13-12q13
C;Superfamily: integrin alpha-2b chain
C;Keywords: glycoprotein; skeletal muscle; transmembrane protein
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-1137/Product: integrin alpha-7 chain, long splice form #status predicted <MATL>
F;34-657,733-1137/Product: integrin alpha-7 chain, short splice form #status predicte
F;1038-1056/Domain: transmembrane #status predicted <TRM>
F;86,742,945,981,1001/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.2%; Score 8; DB 2; Length 1137;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LLVLALLV 14
||||||||
Db 1049 LLVLALLV 1056

RESULT 14
B53283
major cat allergen Fel d I beta chain - cat (fragment)
C;Species: Felis silvestris catus (domestic cat)
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 12-May-1994
C;Accession: B53283
R;Duffort, O.A.; Carreira, J.; Nitti, G.; Polo, F.; Lombardero, M.
Mol. Immunol. 28, 301-309, 1991
A;Title: Studies on the biochemical structure of the major cat allergen Felis domesti
A;Reference number: A53283; MUID:91287714
A;Accession: B53283
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <DUF>

Query Match 6.3%; Score 7; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 PIFYDVF 33
|||||||
Db 8 PIFYDVF 14

RESULT 15
T20911
hypothetical protein F14F8.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20911
R;Lloyd, C.
submitted to the EMBL Data Library, March 1997

A;Reference number: Z19346
A;Accession: T20911
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-73 <WIL>
A;Cross-references: EMBL:Z92782; PIDN:CAB07183.1; GSPDB:GN00023; CESP:F14F8.8
A;Experimental source: clone F14F8
C;Genetics:
A;Gene: CESP:F14F8.8
A;Map position: 5
A;Introns: 16/3; 46/3

Query Match 6.3%; Score 7; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       7 LLVLALL 13
            |||||||
Db       5 LLVLALL 11
```

Search completed: May 7, 2002, 12:18:15
Job time: 111 sec